EPJ manuscript No.
(will be inserted by the editor)

A puzzle in DNA biophysics

Ralf Bundschuh[1] and Ulrich Gerland[2]

[1] Department of Physics, The Ohio State University, 191 W Woodruff Avenue, Columbus, Ohio 43210-1117, USA
[2] Arnold-Sommerfeld Center for Theoretical Physics and Center for Nanoscience (CeNS), LMU München, Theresienstrasse 37, 80333 München, Germany

January 30, 2006

Abstract. In this issue, Lee *et al.* report the experimental temperature-dependence of the unzipping force for two natural DNA sequences. For both sequences, the curves show an anomaly at temperatures around 40^oC. In this brief contribution, we stress that the anomaly is not easily explained within the established theoretical models for the biophysics of DNA. As this puzzle questions our basic understanding of DNA, it must be resolved, most likely by a combination of additional experiments and new theoretical work.

PACS. 87.14Gg DNA, RNA

DNA does not only store the genetic information, it is also the physical substrate on which this information is processed by transcription factors, polymerases, and other proteins [1]. To understand the information processing quantitatively, as well as for applications in nanotechnology [2], it is important that we characterize the biophysics of DNA in detail. Hence, an experimental hint that our current understanding might be incomplete should trigger a deeper investigation. Here, we argue that such an investigation is prompted by the results of Lee *et al.* reported in this issue.

These authors measured the critical force $f_c(T)$ required for the unzipping of DNA as a function of the ambient temperature T. They determined the $f_c(T)$ curves for two natural sequences with different CG content (53 % versus 40 %), but the same length of 2000 base pairs. The overall shape of the two curves is virtually identical, with only a slight shift towards larger forces for the sequence with higher CG content. This finding suggests that the shape of the curve is not related to specific properties of a particular sequence, but a generic property of DNA. It comes as a surprise then that this shape differs from that predicted by the current DNA models: while $f_c(T)$ closely follows the theoretical curve for temperatures below $\sim 38^o$C, it then displays an abrupt reduction of the critical force by about $4pN$ in a temperature interval of about 2 degrees, see Fig. 1 (diamonds and upward triangles). Before we discuss these findings, it is important to review the experimental procedure of Lee *et al.* as well as the assumptions underlying the theoretical calculation of $f_c(T)$.

Experimental critical force.— By definition, f_c is the force at which a closed state where (almost) all base pairs are formed has equal free energy as an open state where (almost) all basepairs are broken. For the experiment, this definition demands that one observes multiple reversible conformational transitions between the closed and open states: f_c is then the force where the time fraction spent in the two states are equal. This is indeed feasible for short molecules such as a single RNA hairpin [3]. In principle, the two long DNA strands in the experiment of Lee *et al.* could also rezip after opening, since they never completely separate. However, rebinding is so slow that the unzipping becomes effectively irreversible during the timescale of the experiment. For this reason, Lee *et al.* used an operational definition for the critical force: many identical DNA's are simultaneously held at a fixed force (applied with the help of magnetic beads) for a certain amount of time τ, and the force value at which half of the molecules yield during this time is taken as a proxy for f_c. Clearly, this operational definition will yield a value that depends on the time τ, as discussed also by S. Cocco in this issue.

Theoretical assumptions.— The simplest model used by Lee *et al.* reduces DNA unzipping to a two-state process. The difference in Gibbs free energy per base pair, ΔG, between the completely zipped state and the completely unzipped state is assumed to depend on temperature via $\Delta G(T) = \Delta H - T\Delta S$, with a temperature-independent enthalpy ΔH and entropy ΔS. Here, ΔH and ΔS are obtained by averaging tabulated values over the sequence composition for the sequence at hand. This amounts to replacing the heterogeneous DNA by an effective homopolymer model. At the critical force $f_c(T)$, the average free energy difference ΔG is exactly compensated by the gain in stretching energy, $2g_u(f)$ for the opening of a single basepair, i.e. $2g_u(f_c) = -\Delta G$. Lee *et al.* obtained $g_u(f)$ from an independent measurement. The resulting theoretical $f_c(T)$ curves are reproduced in Fig. 1 (dashed and dotted lines). As noted above, these curves agree with the experimental data only up to about 38^oC. In order to

test whether the difference above 38^{o}C might be due to the neglect of sequence heterogeneity, Lee *et al.* also performed Monte Carlo simulations where the fully zipped DNA is opened base pair by base pair, with Arrhenius rates using the sequence dependent free energies for DNA base pairing [4] and the same operational definition of the critical force as in the experiment. However, Lee *et al.* found that the simulation results are almost identical with the prediction of the two-state model, over the entire temperature range considered.

As noted already by Lee *et al.*, there are a number of additional assumptions in the theoretical description, which might in principle be responsible for the difference to the experimental $f_c(T)$: (i) the formation of bubbles and alternative secondary structures in the double stranded part of the DNA was neglected, (ii) the possible formation of hairpins in the single-stranded part of the DNA was not included, (iii) the parameters for the enthalpy and entropy changes, ΔH and ΔS, were assumed to be independent of temperature, and (iv) similarly, the free energy $g_u(f)$ of the single strand was assumed to be insensitive to temperature change.

In this contribution, we address the following questions: Could the difference between theory and experiment be due to one of the above simplifying assumptions? Or could the observed anomaly in $f_c(T)$ be caused by the operational definition of the critical force?

Hairpins, bubbles, and alternative structures.— First, we consider the effect of more complex base pairing patterns on the critical force. To this end, we adapted our previously described model for the force-induced denaturation of RNA molecules [5,6] to the case of DNA by using the free energy parameters for DNA, corrected for the experimental salt concentration of 0.15 M [7]. This model incorporates all possible secondary structures of the DNA sequence within the single- and double-stranded part and assumes these structures are in quasi-equilibrium during the unzipping process (see below for non-equilibrium effects). To describe the stretching free energy of the single strand, we use the freely jointed chain model which yields

$$g_u(f) = -kT\frac{b}{\ell_P}\log\frac{\sinh(f\ell_P/kT)}{f\ell_P/kT}\,, \qquad (1)$$

where $\ell_P = 2.1\,\mathrm{nm}$ is the persistence length and $b = 0.7\,\mathrm{nm}$ is the bare base to base distance of single stranded DNA. We apply this model to the same two sequences as in the experiment (we use a hairpin molecule containing the first or last 2100 base pairs of λ-DNA) and determine the critical force f_c as the force at which more than 50% of the molecules in the thermodynamic ensemble have opened at least 2000 base pairs (we verified that other definitions yield very similar results). The resulting critical forces for the two sequences are shown as circles and squares in Fig. 1. We see that even though all possible secondary structures are included, $f_c(T)$ is not significantly different from the dashed and dotted lines obtained from the simple two-state model. Thus, neither the formation of bubbles in the double-stranded DNA nor the formation of secondary structures in the single-stranded DNA contributes to the deviations from the two-state model predictions above 38^{o}C.

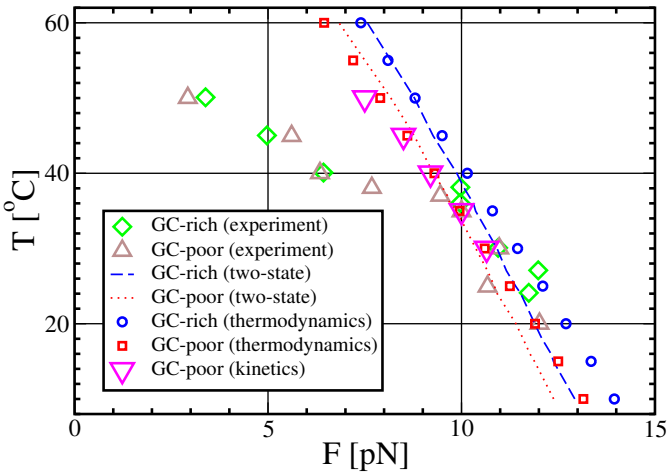

Fig. 1. Unzipping phase diagram of DNA. The diagram shows the critical force for unzipping DNA at different temperatures for a GC-rich (53% GC content) and a GC-poor (40% GC content) sequence. The diamonds and upward triangles represent experimental data from Lee *et al.* in this issue (with permission). The lines represent the two-state model from Lee *et al.* The circles and squares represent the results of a full thermodynamic study that includes all possible base pairing structures such as bubbles in the double-stranded DNA and hairpins in the single-stranded DNA. The downward triangles represent the results of a kinetic study. All the data agrees very well in the temperature range between 20°C and 38°C. At higher temperatures the experimental data shows a pronounced drop in the critical force whereas all theoretical models predict a continuous change of the critical force with temperature while remaining consistent with each other.

Non-equilibrium effects.— Next, we want to discuss the importance of the experimental protocol, which could possibly probe kinetic effects rather than the true thermodynamic critical force (see also the contribution by S. Cocco in this issue). To this end, we performed a full kinetic simulation using the approach outlined in [8] but with realistic stacking energies and including the force-dependent free energy (1) of the freely jointed chain. In order to stay close to the experimental protocol, we started with a completely zipped hairpin at a force below the critical force and simulated the base pairing kinetics at a fixed force for 10^9 elementary base pairing kinetics units which should correspond roughly to the experimental 15 minutes of real time, given that the timescale for an individual base pairing move is on the order of a microsecond. Then, we increased the force in steps of $0.5pN$. In order to make the simulation feasible, we had to constrain ourselves to the first 200 bases of the GC-rich end of the λ-DNA sequence (which has a GC-content of 43% similar to the first 2000 bases of the GC-poor end). We stop the simulation when the completely unzipped state is reached and record the apparent critical force to be the force at which this happens. We repeated each simulation 10 times. For all temperatures either the apparent critical forces agreed among all

10 trials or they were within two consecutive force steps. Fig. 1 shows the critical forces averaged over all 10 trials as the downward triangles. The obtained critical forces are similar to the thermodynamic critical forces. Most importantly, this data shows no force jump around 38^oC, but changes smoothly with temperature. Hence, we conclude that non-equilibrium effects are unlikely the cause of the anomaly.

Temperature dependence of ΔH and ΔS.— Since the enthalpy and entropy changes are usually determined by a linear fit to melting curves in the vicinity of the melting temperature and the regime of the linear approximation is limited, it is clear that ΔH and ΔS will in general be temperature dependent. However, ΔH and ΔS can be expected to vary smoothly with temperature, so that the temperature variation is also unlikely the explanation for the anomaly in the experimental $f_c(T)$ curve. This conclusion is supported by a closer consideration of how these values are obtained experimentally. There are different values of ΔH and ΔS for each of the ten possible stacks of two base pairs. The actual sequence of the molecule then dictates how these sequence dependent values are averaged to obtain the total ΔH and ΔS. The values for the ten possible stacks in turn are determined by melting short DNA double strands of different sequences and measuring the melting temperatures. Since these values are typically used to make quantitative predictions about the structure of a DNA molecule at physiological temperatures, special attention is paid to obtaining reliable values within this temperature range, e.g. by choosing sequences that melt around 37^oC. Specifically, the stacking parameters used by Mfold [9] and us have been constructed to be valid at 37^oC [7]. Thus, the ΔH and ΔS values should be most correct precisely where the deviations between theory and experiment are seen and only start to deviate from these values as the temperature significantly deviates from 37^oC. Lee *et al.* used parameters that were determined from melting experiments in the vicinity of 75^oC [4]. Thus, their values of ΔH and ΔS should be most appropriate around that temperature. The fact, that their values nevertheless quantitatively agree with the experiments in the temperature range of 20^oC to 38^oC suggests that these parameters are not strongly temperature dependent. Indeed, it is unlikely that the parameters are correct between 20^oC and 38^oC as well as around 75^oC but fail in the intermediate regime.

Discussion.— So what could be the origin of the sudden jump in the melting curve? Until now, we have not discussed a possible non-trivial temperature dependence in the mechanical response of the single-stranded DNA, i.e. in the free energy $g_u(f)$ of the single strand. For instance, might there be a jump in the effective bond angle or the persistence length maybe due to a transition in the structure of the single-stranded molecule like the one observed recently [10] in the absence of a force? This question will have to be addressed experimentally, possibly in combination with atomistic modeling. A direct way to probe this experimentally would be to remeasure $g_u(f)$ at different temperatures, using the same approach as Lee *et al.* did for their measurement at 25^oC. Of course, the sudden jump may also be caused by a completely different effect not even considered here. In any case, we feel that this puzzle, which questions our basic understanding of DNA biophysics, must be resolved.

Acknowledgments.— We acknowledge stimulating discussions with Y. Kafri, D. Nelson, and M. Prentiss that initiated the first steps of this work. This work was funded by the Petroleum Research Fund of the ACS through Grant No. 42555-G9 (to RB), by the Ohio Supercomputer Center (to RB), and by an Emmy Noether Grant of the DFG (to UG).

References

1. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell*, (Garland, 2002)
2. N. C. Seeman, Nature **421**, 427 (2003)
3. J. Liphardt, B. Onoa, S. B. Smith, J. Tinoco, Ignacio, and C. Bustamante, Science **292**, 733 (2001)
4. R. Blake and S. Delcourt, Nucleic Acids Res. **26**, 3323 (1998)
5. U. Gerland, R. Bundschuh, and T. Hwa, Biophys. J. **81**, 1324 (2001)
6. U. Gerland, R. Bundschuh, and T. Hwa, Biophys. J. **84**, 2831 (2003)
7. J. SantaLucia, Jr., Proc. Natl. Acad. Sci. USA **95**, 1460 (1998)
8. R. Bundschuh and U. Gerland, Phys. Rev. Lett. **95**, 208104 (2005)
9. M. Zuker, Nucleic Acids Res. **31**, 3406 (2003)
10. J. Zhou, S. Gregurick, S. Krueger, and F. Schwarz, Biophys. J. **90**, 544 (2006)