Bubbles in DNA melting

2009 J. Phys.: Condens. Matter 21 034102

(http://iopscience.iop.org/0953-8984/21/3/034102)

IOP Publishing Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **21** (2009) 034102 (9pp) doi:10.1088/0953-8984/21/3/034102

Bubbles in DNA melting

Rodrigo Gonzalez, Yan Zeng, Vassili Ivanov and Giovanni Zocchi

Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, CA 90095-1547, USA

Received 9 June 2008
Published 17 December 2008
Online at stacks.iop.org/JPhysCM/21/034102

Abstract

This paper presents a review of largely our own work on the DNA melting transition, and some new measurements of the elastic energy of sharp bends in single-stranded DNA and RNA. The purpose is to present the point of view that studying the transition of intermediate size oligomers leads to valuable tests of the models, in particular the ingredients most important for a reduced-degrees-of-freedom description, such as the different role of base pairing and base stacking. We make the case that, with intermediate size oligomers, one can actually measure the bubble length, which exhibits a more interesting behavior than the fraction of dissociated bases alone. Here is where more work seems necessary, both on the experimental and the modeling side, to understand the differences between theory and experiments. We summarize our previous results on the cooperativity parameters, which suggest that the transition is never exactly two-state no matter how short the molecule, or in other words the nucleation size for bubbles opening at the ends of the molecule is essentially 1 base pair. We briefly discuss our own modification of the nearest-neighbor model which treats pairing and stacking separately, as a way to fit the experimental melting profiles in this intermediate length regime. Finally we go on to present some new measurements on the stability of DNA and RNA hairpins with very short loops.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The ordered helical structure of double-stranded (ds) DNA and long polypeptide helices (such as poly-lysine) melts at 'high' temperature into a more disordered state. It was recognized early on that, for long molecules, the transition would exhibit coexistence of helical and non-helical segments in the same molecule [1], so for a statistical mechanics description the central problem becomes calculating the statistical weight of a given non-helical segment with respect to the weight of the helical segment (which in the following we take equal to 1, thus counting energies and entropies with respect to the reference ordered state). The statistical weights are fundamentally different for the case of the polypeptide helix and ds DNA, because the non-helical DNA segment is actually a loop (except if it is located at one end of the molecule), and because binding in the DNA is between nearest neighbors (both across the two strands and along the same strand, corresponding to base pairing and base stacking) whereas in the polypeptide α-helix each aminoacid hydrogen bonds to the fourth-nearest neighbor.

In the case of DNA, the statistical weight of an ss segment (bubble) is built up from the energy of opening the bases and the entropy of the loop. How these are evaluated depends on the problem at hand. The approach favored by those interested in the properties of the transition in the thermodynamic limit of an infinite molecule is to take a simple form for the energy (namely a fixed energy per open bp) and calculate the entropy of the loop from fundamental polymer theory [2–4]; this is a delicate and interesting problem (largely because of excluded-volume effects) but it is now essentially understood. At the other extreme, those interested in the melting profiles of short oligomers take the approach of writing the free energy of an ss segment as the sum of the free energies of individual 'dimers', two consecutive base pairs forming a dimer. This is known as the nearest-neighbor (NN) model [5]; the reason for considering dimers is that, experimentally, the free energy of opening a bp also depends on the identity of the next bp; this is due to the stacking interaction between neighboring bases on the same strand. We may say that in this approach the energy is modeled in detail but the entropy is simplified (scaling linearly with the length of the loop). However, for sufficiently short oligomers there are no loops and the NN model gives an accurate description. Because of the directionality of the DNA strands, there are ten different dimers; melting profiles of short ($<$10 bp) oligomers can be analyzed in terms of a

two-states model to yield an energy and an entropy parameter, so from the melting profiles of different sequences the free energy parameters for all ten dimers have been measured [5]. The challenge is then whether the NN model thus constructed can be extended to accurately describe the melting transition of longer sequences. One approach is to introduce into the bare NN model fixed free energy costs for opening a loop or opening one end; this has the advantage of simplicity but it does not cure the problem of the wrong scaling of the entropy with loop length. However, it is also possible to incorporate the PS loops with the correct entropy into the NN model [6–8]. So do these models accurately describe the transition for medium size and long molecules?

If one is only interested in the melting temperature (the midpoint of the transition determined by a particular melting profile, usually a UV absorption curve), then the online servers based on the various incarnations of the NN model above generally reproduce the data with a level of accuracy comparable to the experiments. Note that the question is the accuracy of the model using a fixed set of parameters, usually optimized on the melting profiles of short oligomers, as mentioned above. Because of the large number of parameters (22 in the NN model), it is typically possible to reproduce any given melting profile almost exactly with small adjustments of the parameters.

However, if one is interested in an accurate description of intermediate states (i.e. bubble states), then the model generally falls short of reproducing the experimental data [8–10]. Specifically, the fine structure of the melting profiles of kb long molecules (conveniently displayed as a series of peaks in the $\mathrm{d}f/\mathrm{d}T$ versus T profile, where f is the UV absorption signal reporting of the fractional length of the molecule in the ss state) and the occurrence of bubble states in medium size oligomers (20–60 bp), which we discuss below, as well as the influence of mismatches on the bubble states, is not completely captured by the models. Whether this is because the NN model parameters are, in fact, not yet fully optimized, or because additional detail should be introduced in the model (such as, for instance, the helical structure of the ground state), is not clear at the moment [8]. In the following, we summarize our experimental work on oligomers of intermediate length, which provide stringent tests for the models, and focus on the question of the most relevant degrees of freedom for a coarse-grained description. We present the view, borne from the experimental melting profiles of intermediate size oligomers, that it is advantageous to keep base pairing and base stacking degrees of freedom separate (instead of combining them into effective free energies). Then we go on to report some new experimental results on the elastic energy of sharp bends in ss DNA and RNA.

The overriding theme of this issue of the journal being the physico-chemical understanding of DNA which is relevant for biology, understanding bubble states in long molecules is important in genomics (see the article by Bishop and Rasmussen in this issue), while intermediate length oligomers and their bubble states are important in biotechnology applications such as PCR primers (including mismatched primers which are used for directed mutagenesis) and silencing RNAs. Sharp bends in single strands of DNA and RNA relate

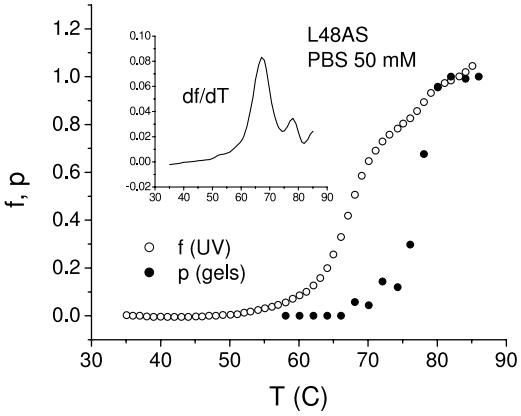

Figure 1. Melting profile for a 48 bp long DNA oligomer (codenamed L48AS [11]); this same oligomer is represented in the hairpin state in figure 3. The open circles are the UV absorption curve $f(T)$, the filled circles represent the strand dissociation curve $p(T)$ obtained with the quenching method. $f(T)$ is normalized such that $f = 1$ for $p = 1$. The inset shows $\mathrm{d}f/\mathrm{d}T$.

to the question of tertiary structure formation, presently a lively topic in biotechnology with the accelerating development of new aptamers.

Note that we have not included the Peyrard–Bishop–Dauxois model, which is very successful in describing many aspects of the DNA melting profiles, in our discussion of the models, because it is discussed in other papers in this issue.

2. Why study intermediate length oligomers?

Short oligomers (∼10 bp or less) melt through a transition which is very close to two-state, so one cannot study bubbles. Melting profiles of long (∼kb) DNA reveal the existence of many bubble states (the $\mathrm{d}f/\mathrm{d}T$ versus T curve shows many peaks) but this fine structure is complicated: for instance, it is difficult to assign a particular peak to the melting of a specific segment. Intermediate length oligomers (∼20–100 bp) display relatively simple melting profiles (the $\mathrm{d}f/\mathrm{d}T$ versus T curve has one or two peaks (figure 1)) but bubble states are nonetheless prominent. This is actually not obvious, because spectroscopic (say, UV absorption) melting profiles $f(T)$ often are rather featureless sigmoidal curves (figure 2), i.e. $\mathrm{d}f/\mathrm{d}T$ has essentially one peak, which can equally well be interpreted in terms of a two-state transition. The reason is that spectroscopic methods such as UV absorption report on the fraction of open bp without discriminating between partially and totally dissociated strands. The midpoint ($f = 1/2$) of the normalized melting curve can then in principle correspond to a state where half the molecules in the sample are completely dissociated and half are completely closed, or a state where all molecules are halfway open, or anything in between. Clearly what is needed is an independent dissociation curve $p(T)$ which counts the fraction of completely dissociated molecules.

A method to measure $p(T)$ was introduced in [11, 12] and is summarized below. To reiterate, there are three distinct length regimes of interest: short oligomers, which melt in a transition close to two-state, intermediate length oligomers, for

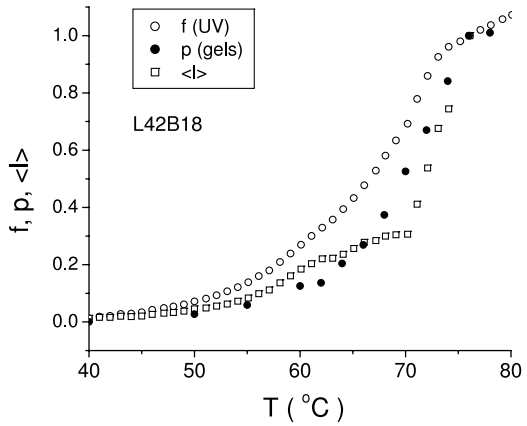

Figure 2. Melting profiles for the sequence L42B18 (reproduced from [9], copyright 2004 Elsevier), which consists of an AT-rich middle region of length $B = 18$ bp ('bubble forming region') clamped at the ends by GC-rich regions, for a total length of the molecule $L = 42$ bp. The open circles represent the fraction of open base pairs f (from the UV absorption measurements); the filled circles represent the fraction of dissociated molecules p (from the quenching method) and the squares represent the average length of the bubble ℓ (supposing there is only one bubble), which is calculated from equation (1).

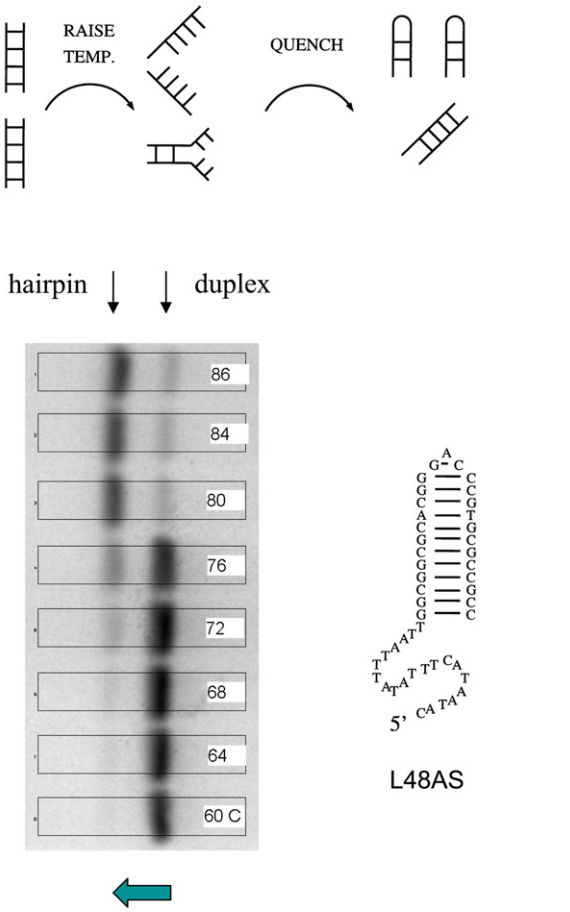

Figure 3. The principle of the quenching method to determine the fraction of completely dissociated molecules $p(T)$. Partially self-complementary sequences are annealed into the duplex state (the ground state). The temperature is then raised to an intermediate value within the transition region, followed by a quench to ∼0 °C. Molecules which were completely dissociated before the quench form hairpins immediately after the quench, while partially open molecules close back as duplexes. Hairpins (hp) and duplexes (ds) in the quenched sample are separated by gel electrophoresis and their relative populations measured from the intensities of the corresponding bands on the gel. The numbers on the lanes give the temperature to which the sample was heated before the quench. The relative population of hairpins and duplexes reports on the relative populations of completely dissociated and partially opened molecules at that temperature (see equation (2)). Also shown is the hairpin structure for the sequence L48AS.

which the transition is not two-state but is still dependent on oligomer concentration (i.e. there are finite size effects), and long molecules where the transition is essentially independent of concentration. The intermediate regime is interesting because, first of all, one can measure two meaningful 'melting curves' instead of one (namely, the fraction of dissociated bases $f(T)$ and the fraction of dissociated molecules $p(T)$); this enables more stringent tests of the models [13, 14] and raises the possibility of further optimizing parameters of the thermodynamic models obtained from short oligomers [8]. Second, in the case where only one bubble is present (which can be achieved by playing with the sequence [12]) one can measure the length of the bubble $\ell(T)$, because

$$p + (1 - p)\ell = f \Rightarrow \ell(T) = \frac{f - p}{1 - p}. \qquad (1)$$

So the main point in this intermediate regime is that, whether doing experiments [12] or computations [14], one has to deal with averages over two different ensembles: the ensemble of all molecules and the sub-ensemble of the non-dissociated molecules [14].

The dissociation curve $p(T)$ can be measured as follows. One chooses sequences which are partially self-complementary, so they can form both duplexes and hairpins. One works in a regime of concentrations where the duplex is more stable but the hairpin, once formed, is sufficiently long-lived. The sample is initially annealed into the duplex state, then brought to a temperature T within the transition region, and finally quenched to ∼0 °C. Strands which were dissociated at temperature T form hairpins after the quench, while partially open molecules close again as duplexes (figure 3). Hairpins and duplexes in the quenched sample are separated by gel electrophoresis, the relative amount of hairpins reporting on the fraction of dissociated strands at the temperature T before the quench. Thus $p(T)$ is obtained from

$$p(T) = \frac{\mathrm{hp}}{\mathrm{hp} + \mathrm{ds}} \qquad (2)$$

where hp is the amount of hairpins, ds the amount of duplexes and T is the temperature before the quench. For a detailed description of experimental corrections to (2) and controls, we direct the reader to the published work [9, 12]. Combining the dissociation curve $p(T)$ obtained with this quenching method with the traditional melting curve $f(T)$ obtained from UV absorption (which reports on the fraction of unpaired bases), one obtains a picture of the melting transition for intermediate

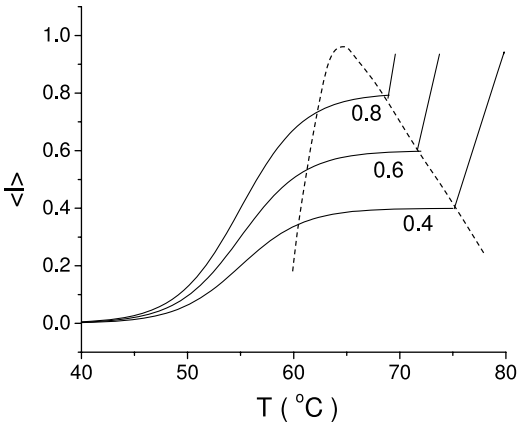

Figure 4. Schematic representation of the ℓ–T curves for different ratios B/L (these are *not* experimental data: the curves are a schematic representation of the trend in the experimental melting profiles of figure 2 in [9]); ℓ is the length of the bubble, B the length of the central 'bubble forming region' in the sequence and L the length of the sequence. For increasing B/L, the plateau in the $\ell(T)$ curve, which occurs at $\ell \approx B/L$, contracts in size (see [9]). The numbers next to the curves represent the corresponding value of B/L. The dotted line encompasses the region where there is coexistence of bubble states and dissociated states.

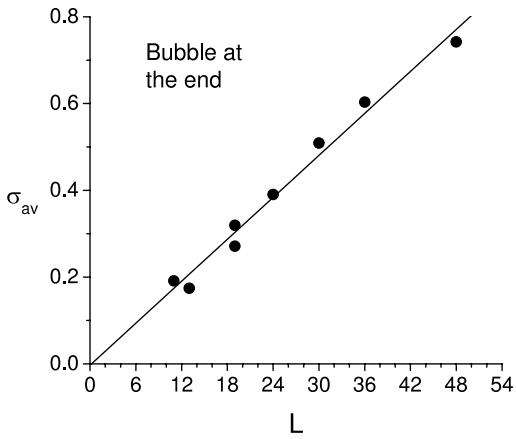

Figure 5. $\sigma_{\rm av}$ is the average of $\sigma = f - p$ (equation (3)) over the transition region, and is a measure of the statistical weight of bubble states. $\sigma_{\rm av}$ is plotted versus the length of the sequence L for eight sequences which form bubbles at the ends (reproduced from [9], copyright 2004 Elsevier). Extrapolation of the data (the straight line is a linear fit) suggests that the transition becomes strictly two-state ($\sigma_{\rm av} = 0$) only for $L \approx 1$.

length oligomers as in figure 2. The $f(T)$ profile alone is rather featureless, but $p(T) < f(T)$ in the transition region reveals the existence of bubble states, and the $\ell(T)$ profile calculated from (1) has the striking feature of a plateau. This sequence is such (an AT-rich stretch of 18 bp clamped at both ends by 12 bp long GC-rich segments; the nomenclature L42B18 refers to a molecule of length 42 bp with a 'bubble forming' central region of length 18 bp) that a bubble opens 'in the middle'; the plateau in $\ell(T)$ is the signature of the coexistence of bubble states and dissociated states in the sample.

We wish to briefly pursue the perhaps superficial analogy with the isotherms in the P–V diagram for a liquid–gas transition. We observe that in our case the width of the plateau in $\ell(T)$ decreases for increasing length of the central AT-rich region (for increasing B with the nomenclature above) (figure 2 of [9]); this is seen in the experiments and must be the case because the dissociation curve $p(T)$ becomes increasingly steeper for increasing B; if $p(T)$ is infinitely steep, $\ell(T) = f(T)$ for $0 \leqslant f \leqslant 1$ and there is no plateau. So we come to a picture of the ℓ–T curves for different B which qualitatively looks like figure 4, where the dotted line denotes the boundary of coexistence of bubble states and dissociated states. In the figure, we indicate the existence of a critical point for $B/L \to 1$; this limit corresponds to a homogeneous sequence (uniformly AT-rich). In this picture a homogeneous sequence has, in the thermodynamic limit, a continuous transition, characterized by the absence of coexistence between bubble and dissociated states, or, which is the same, characterized by a divergent bubble length. This is qualitatively in agreement with the conclusions of [15].

However, the entire above paragraph is conjecture. Coming back to comparing models and experiments, the PBD model reproduces the melting profiles of intermediate length oligomers very well [13] (note, however, that some of the conclusions have been criticized in [14]), without retouching the parameters; the NN model can probably also describe these profiles accurately but it needs some further parameter optimization (within the current uncertainty in parameter values [8]). We now address in more detail certain points which are raised by the melting profiles of intermediate length oligomers.

3. The cooperativity parameters

The quantity

$$\sigma = f(T) - p(T) \tag{3}$$

represents the fraction of bases which participate in a bubble state. If we plot the average value of σ through the transition region, $\sigma_{\rm av}$, for different lengths L of the molecule, we obtain the graph of figure 5 [9]. The longer of these sequences ($L = 24$ and up) are designed to favor opening from one end (that end is AT-rich). It is not possible to use the quenching method for very short sequences (<10 bp), so we do not know how this curve looks like for $L < 10$, but if we draw a line through the data we get the suggestive result that the fit extrapolates to the origin. That is, if we believe that $\sigma_{\rm av}$ goes to zero linearly with L: $\sigma_{\rm av} \propto (L - L_0)$, then we find $L_0 \approx 0$ from the experiments. This says that the cooperativity parameter or extra free energy cost of opening the base pair at the end of the molecule is essentially zero. Therefore the transition is never exactly two-state, no matter how short the molecule.

We find this result plausible, because the end of the molecule is already a 'defect', and known to be 'fraying', so there is no compelling physical reason to expect a finite nucleation size for bubbles opening at the ends of the molecule. However, our measurements do of course not exclude that the $\sigma_{\rm av}$ graph (figure 5) might turn sharply downwards for $L < 10$ and intersect the abscissa at a finite L.

On the other hand, if we look at bubbles opening in the middle of the molecule (again using an opportune

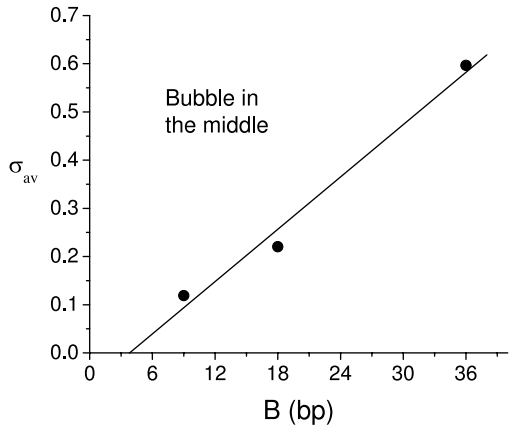

Figure 6. The same plot as figure 5 for the three sequences from [9] which form a bubble in the middle. σ_{av} is plotted versus the length of the bubble forming region B.

choice of sequences [9]) we find instead a finite nucleation size (figure 6). In figure 6, we have displayed σ_{av} versus B (the length of the bubble forming AT-rich region in the sequence): the straight line intersects the abscissa at $L_0 \approx 4$; however, if we plot instead σ_{av} versus L (the total length of the molecule including the GC end-clamps) the plot gets shifted to the right by 24 bp, i.e. now the line intersects the abscissa at $L_0 \approx 28$. The true value of the nucleation size (which is related to the cooperativity parameter for bubble opening in the middle, which is certainly not zero) lies in between these extremes, and could be obtained from such plots by generating more data with varying length of the GC end-clamps.

4. Base pairing and base stacking as independent degrees of freedom

The melting profiles of oligomers obtained by UV absorption measurements do not display constant baselines before and after the transition. Instead, the UV absorption signal $f(T)$ keeps rising after the critical temperature of strand dissociation (figure 7). As is well known, this is due to unstacking in the single strands: $f(T)$ contains contributions from both unpairing and unstacking, i.e. in a linear approximation

$$f(T) = \alpha \langle \text{unpaired} \rangle + \beta \langle \text{unstacked} \rangle . \qquad (4)$$

So there are really two transitions: an unpairing and an unstacking transition, and they are partially overlapping in temperature. Furthermore, they are not independent because, for instance, if two adjacent bases are unstacked at least one of them must be unpaired also. Clearly for an accurate description of the melting transition we must include pairing and stacking as separate degrees of freedom, as they show up individually in the measurements. This is not a small effect (see figures 7, 4 and 5 in [10]): stacking energies are comparable to pairing energies, and while the midpoint of the unstacking transition is generally considerably above the strand dissociation temperature, this is a very broad, non-cooperative transition (see below) so it reaches down to lower temperatures and overlaps with the unpairing transition. In short, in a melting curve such as shown in figure 7 there is a

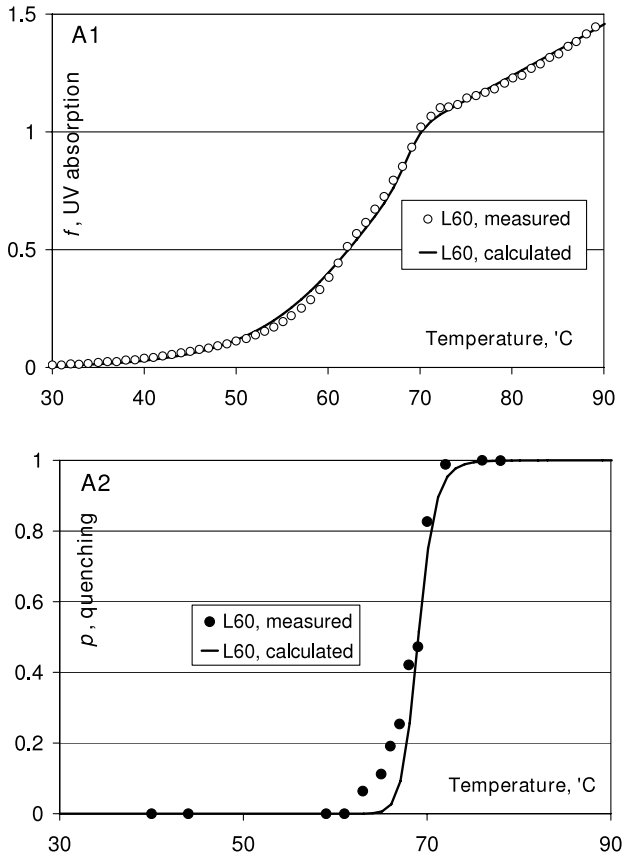

Figure 7. A1: normalized $f(T)$ profile (from UV absorption measurements); A2: dissociation curves p (from the quenching method) for the sequence L60 [17]. The experiments are the circles; the solid lines are the model of section 4 [17].

measurable contribution of unstacking to the $f(T)$ curve even in the region $f < 1$ (we are normalizing $f(T)$ curves such that $f = 1$ corresponds to complete strand dissociation, i.e. $f = 1$ coincides with $p = 1$).

In the NN model pairing and stacking are lumped together into the effective free energies of the dimers. In the PBD model, stacking is instead introduced separately from pairing in the nonlinear terms of the Hamiltonian. Whether this Hamiltonian accurately describes the separate contributions of unpairing and unstacking to the melting profile $f(T)$ is an open question because the PBD model has not been used so far in conjunction with (4) to explore the entire melting profile $f(T)$ of oligomers including the region $f > 1$. However, it is also possible to revisit the NN model and keep pairing and stacking degrees of freedom separate. Before we discuss below one realization of this approach, we draw attention to certain experimental facts about the unstacking transition.

As mentioned above, the transition is broad and the midpoint often lies about or above 100 °C; however, it turns out that by lowering the pH this transition can be moved to lower temperatures, which allows us to measure the whole melting profile [16]. As shown in figure 8, unstacking is a broad transition (compared to unpairing) and essentially non-cooperative, in the sense that it is well described by a simple Ising model of spins with nearest-neighbor coupling.

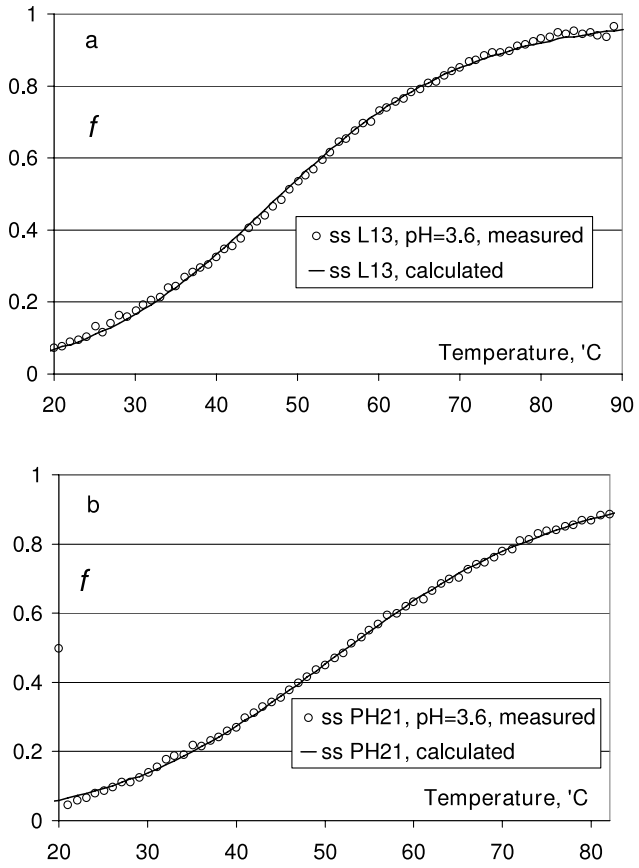

Figure 8. Melting curves obtained from UV absorption for the ss oligomers L13 (a) and PH21 (b). Circles are the experimental data; solid lines are fits with the Ising model. The data were obtained at pH = 3.6 in order to lower the midpoint of the unstacking transition. L13 is almost self-complementary, but at this pH the ds structure is not stable at room temperature; PH21 is non-self-complementary, so there is no ds structure at any pH. The unstacking transition is much broader in temperature compared to double helix melting, and is well described by the Ising model.

In light of the above, we modified the NN model by allowing unpaired bases to be either stacked or unstacked [17]. This approach has three advantages: (1) it includes both transitions (unpairing and unstacking) and so can describe the melting profiles in the entire range (including $f > 1$, i.e. above the strand dissociation temperature); (2) it allows us to introduce a local mechanism for cooperativity based on a simple geometric constraint, thus avoiding introducing an extra 'cooperativity parameter' (free energy penalty for opening a bubble) in the model (this aspect is similar in the PBD model); (3) it accounts for the residual temperature dependence of the dimer free energies of the original NN model.

There are, however, two main disadvantages. One is that the thermodynamic parameters for stacking are less well known than the combined pairing and stacking parameters of the NN model, precisely because of the difficulty of separating the two contributions. The second disadvantage is the increased number of parameters: there are in principle 16 different stackings and 2 pairings (and no extra cooperativity parameter in our formulation [17]), i.e. 18 parameters versus

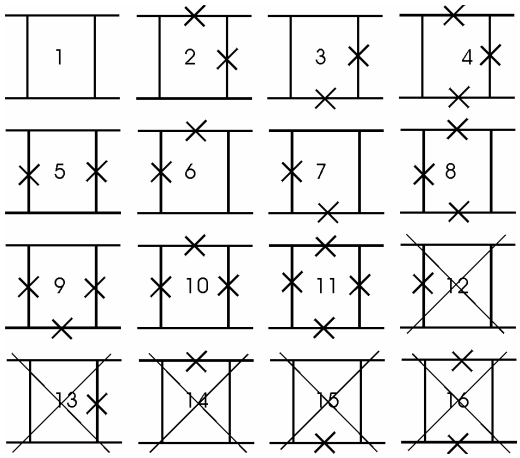

Figure 9. Each NN dimer has two pairings (vertical lines) and two stackings (horizontal lines). The broken bonds are crossed. The horizontal lines represent the strands. There are 16 states of the dimer; admissible states of the model of section 4 are the states from 1 to 11; the states from 12 to 16 are prohibited by the geometric constraints.

10 different dimers plus 1 cooperativity parameter, i.e. a total of 11 parameters, for the NN model.

Nonetheless, the model is useful for understanding in simple terms the mechanism of cooperativity of bubble opening and in general the interplay between pairing and stacking. With the notation that $G_i^{\mathrm{p}} = E_i^{\mathrm{p}} - TS_i^{\mathrm{p}}$ is the free energy of unpairing the i th base, G_i^{st} the free energy of unstacking bases i and $(i+1)$ (of the same strand), $G_i^{\mathrm{st*}}$ the free energy of unstacking the corresponding bases of the complementary strand, the free energy for complete unpairing of the NN dimer is:

$$G_i^{\mathrm{NN}} = (G_i^{\mathrm{p}} + G_{i+1}^{\mathrm{p}})/2 - T\{\ln[1+\exp(-G_i^{\mathrm{st}}/T)] + \ln[1+\exp(-G_i^{\mathrm{st*}}/T)]\} \quad (5)$$

which accounts for the temperature dependence of the thermodynamic parameters (enthalpies and entropies) of the NN model. The partition sum of the system can be written using the NN dimers and in the transfer matrix formalism (as far as pairing is concerned, since two successive dimers share a pairing), but now there are internal states to these dimers because of the stacking. There are in all 16 possible states of the dimer (figure 9) and the cooperativity of bubble opening arises naturally by excluding five of these states, the rule being that you cannot break exactly one pairing in the dimer without breaking at least one stacking, and if both pairings are intact then both stackings are intact too. In terms of the statistical weights

$$U_i^{\mathrm{p}} = \exp\{-G_i^{\mathrm{p}}/T\} \text{ etc} \quad (6)$$

the statistical weight of a given configuration of the NN dimer is then built up using the remaining relevant diagrams in figure 9, for example the statistical weight of the dimer with both pairings open corresponds to diagrams 5, 9, 10 and 11 and is

$$\sqrt{U_i^{\mathrm{p}}}\sqrt{U_{i+1}^{\mathrm{p}}}\left[1+U_i^{\mathrm{st}}\right]\left[1+U_i^{\mathrm{st*}}\right] = \sqrt{U_i^{\mathrm{p}}}\sqrt{U_{i+1}^{\mathrm{p}}} \times \left[1+U_i^{\mathrm{st}}+U_i^{\mathrm{st*}}+U_i^{\mathrm{st}}U_i^{\mathrm{st*}}\right] \quad (7)$$

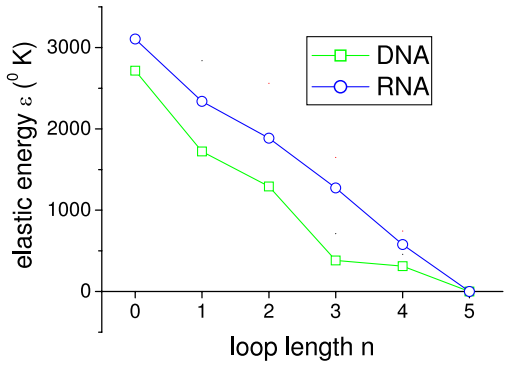

Figure 10. The elastic energy ε as a function of loop length, extracted from the fits to the melting profiles according to equations (8)–(14) in section 5. Squares represent the DNA data and circles the RNA data.

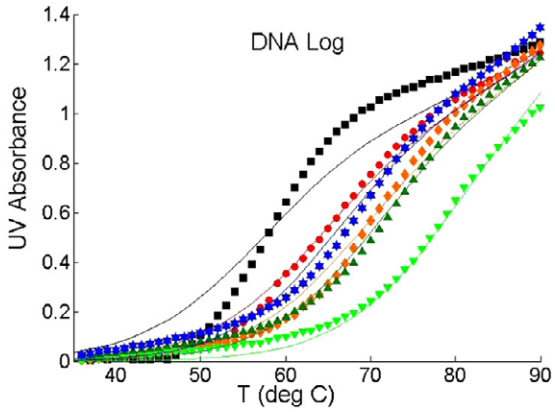

Figure 11. The DNA melting profiles for different loop lengths n and the fits from which the elastic energies of figure 10 were extracted. Squares: $n = 0$; circles: $n = 1$; diamonds: $n = 2$; inverted triangles: $n = 3$; triangles: $n = 4$; stars: $n = 5$.

where the sum in the last square bracket is the sum over the internal states represented by the above diagrams. Figure 7 shows an example of using this model to fit the melting (f) and dissociation (p) curves of an oligomer of intermediate length (60 bp). We used different thermodynamic parameters for pairing for AT and GC, and, for simplicity, the same thermodynamic parameters for all the stackings. At this level, the main interest of the model is to account for the unstacking contribution to the melting curve f, clearly visible in figure 7 where, beyond the temperature where $p = 1$ (all strands dissociated), $f(T)$ keeps rising.

5. Sharp bends in DNA and RNA single strands

In the preceding paragraphs we pursued the point of view that experiments on intermediate length oligomers are informative because they provide stringent tests of the models in the form of distinct melting and dissociation curves, and measurements of bubble length. The contribution of unstacking, evident in these experiments, is important particularly because the interplay of pairing and stacking is responsible for the cooperativity of bubble opening. Another way to look at the cooperativity problem is to ask what is the free energy cost of regions of large curvature ('kinks') of the DNA backbone. This may also be interesting in the context of understanding the different propensity of DNA and RNA to form functional tertiary structures. In fact, we wanted to investigate this latter point, and in the following we present some new experimental results on the energy cost of kinks in ss DNA and RNA.

Our kinks are hairpin structures consisting of a stem and a loop; the stem is a 7 bp fixed sequence, and we vary the length n of the poly-T loop between 0 and 5 bases. The measurement is the melting profile of the hairpin obtained by UV absorption, and we are interested in extracting an 'elastic' energy associated with the loop. More precisely, we are interested in investigating whether there are differences in this elastic energy between DNA and RNA. All measurements are performed under the same conditions, namely 100 mM PBS buffer.

As the loop length n is increased from $n = 0$, we expect the corresponding hairpins to first become more stable, as the destabilizing elastic energy of the loop is reduced for longer loops, and then become once again more unstable, because of the growing difference in loop entropy between the 'closed' and 'open' states of the hairpin. To say the same in terms of rates, a short loop ($n \sim 0$–3) increases the opening rate of the hairpin (and has little effect on the closing rate), while a long loop ($n > 3$–4) decreases the closing rate (with little effect on the opening rate). Indeed, the melting profiles of hairpins show a non-monotonic shift in the midpoint of the transition with increasing loop length n (figure 11). From these melting curves, we want to extract an elastic energy associated with the loop. We are mainly interested in differences between DNA and RNA. Our strategy is then to fit the melting profiles with a reasonably simple model; we want as few parameters as possible and try to keep most of them the same for the DNA and RNA fits, in order to explore possible differences in the loop elastic energy between DNA and RNA.

We use a two-state model for the opening of the stem, because it is so short. We introduce independent unstacking degrees of freedom through an Ising model [16] because it is necessary to fit the $f(T)$ curves in the entire temperature range. We introduce an entropy of the loop with a specified n dependence to account for the destabilizing effects of long loops. Finally, we extract from the fits values of the 'elastic energy' of the loop for different n. The model is defined by the partition sum:

$$Z = 1 + \mathrm{e}^{-\frac{U}{T}+\sigma}\left(1 + \mathrm{e}^{-\frac{E}{T}+S}\right)^{2(N-1)}. \tag{8}$$

With the bracket equal to 1 this is a two-state model for the melting of the stem, U is the energy cost of opening the stem and σ the entropy gain. The bracket expresses the Ising description of unstacking, where N is the number of bp in the stem, E the unstacking energy per base and S the unstacking entropy. We are going to fix these stacking parameters once and for all (to the same values for DNA and RNA) and concentrate on U and σ. With n the number of bases in the loop, we write

$$U = U_0 - \varepsilon(n) \tag{9}$$

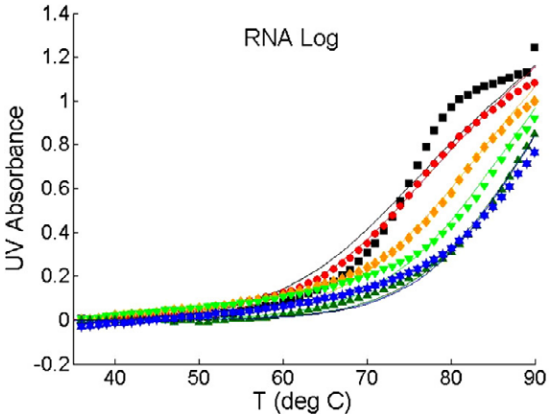

Figure 12. The RNA melting profiles corresponding to the DNA profiles of figure 11. Squares: $n = 0$; circles: $n = 1$; diamonds: $n = 2$; triangles: $n = 3$; stars: $n = 4$; inverted triangles: $n = 5$.

where U_0 is the same for all DNA curves, and a different but again uniform value for the RNA curves. A different U_0 for RNA and DNA is needed to account for the well-known difference in stability between DNA and RNA duplexes. $\varepsilon(n)$ is the 'elastic energy' of the loop which we want to measure, and we expect the properties $\varepsilon(n) \geqslant 0$ and $\varepsilon(n) \to 0$ for n large.

For the entropy σ we take the form

$$\sigma = \sigma_0 + \sigma_1 \ln n \qquad (n \geqslant 1) \tag{10}$$

where σ_0 is the same for all DNA and RNA curves and represents the entropy gained by opening the stem. The σ_1 term represents the entropy gain of the loop when the stem opens. The rationale is that the entropy difference between an unconstrained strand of n bases and a strand of n bases constrained to close in a loop is $\propto \ln n$ (this is a good approximation for large n, but we use it for small n also, which is a weak point of this analysis). In summary, we fix once and for all a form $\sigma = \sigma(n)$ for the entropy, so that the only fitting parameter to fit the melting profiles for different n is $\varepsilon(n)$. The objective is to detect differences in $\varepsilon(n)$ between DNA and RNA.

To complete the description we calculate from (8) the expectation value for the stem to be open:

$$\langle \text{unpairing} \rangle = -T \frac{\partial}{\partial U} \ln Z = \{1 + \exp(U/T - \sigma) \times [1 + \exp(-E/T - S)]^{-2(N-1)}\}^{-1} \tag{11}$$

(this is a number between 0 and 1, where 0 is the closed state and 1 is the open state), and the expectation value for unstackings:

$$\langle \text{unstackings} \rangle = -T \frac{\partial}{\partial E} \ln Z = \langle \text{unpairing} \rangle \frac{2(N-1)}{1 + \exp\{E/T - S\}} \tag{12}$$

(this is a number between 0 and $2(N-1)$) and relate these to the measured melting profile $f(T)$ through

$$f(T) = c\{N\beta \langle \text{unpaired} \rangle + \delta \langle \text{unstacked} \rangle + \gamma\} \tag{13}$$

where c is the oligomer concentration, β, δ parameters describing the change in UV absorption upon unpairing and unstacking, respectively, and γ the absorption of the ground state duplex. By normalizing the experimental profiles so that $f = 0$ below the transition and $f = 1$ at the strand dissociation temperature (beyond which $f > 1$ because of unstacking), and performing all experiments at the same concentration of oligomers, we can rewrite the above relation as

$$f(T) = \langle \text{unpaired} \rangle + \alpha \langle \text{unstacked} \rangle \tag{14}$$

where the number of bp in the stem, N, which is the same for all experiments, has been absorbed in the parameter α.

Using (11), (12) and (14) we fit the melting profiles through the following procedure. Starting with DNA, we use the curve for $n = 5$ (where the contribution from ε is small) to fix E, S, U_0 and $\sigma(5)$ (see equations (8)–(10)), i.e. we assume $\varepsilon(5) \approx 0$. Then we use the $n = 5$, 4, 3 curves (where ε is presumably still small) to find values for σ_0 and σ_1. Now all parameters are fixed except $\varepsilon(n)$, and we re-fit all the curves with these fixed parameters and find $\varepsilon(n)$. Then we do small adjustments of, for example, σ_1, etc, and see whether $\varepsilon(n)$ changes considerably or not, i.e. we test how robust this determination of $\varepsilon(n)$ is with respect to the choice of parameters such as σ_1, etc.

For RNA we keep the same values for E, S, σ_0 and σ_1, find a new value of U_0 and again determine $\varepsilon(n)$.

The results for $\varepsilon(n)$ are shown in figure 10, and the melting curves and fits are displayed in figures 11 and 12. The conclusion is that, from these data, we do not see any significant difference in the behavior of the energy of the loop between DNA and RNA. These measurements could be improved, in particular by moving the transitions towards lower temperatures (by reducing salt concentration, for example), so as to see a bigger part of the transition region, particularly for RNA. For now, this is a null result: no significant difference between DNA and RNA.

References

[1] Poland D and Scheraga H A 1966 *J. Chem. Phys.* **45** 1456–64
[2] Fisher M E 1966 *J. Chem. Phys.* **45** 1469–4
[3] Azbel M Ya 1979 *Phys. Rev.* A **20** 1671
[4] Kafri Y, Mukamel D and Peliti L 2000 *Phys. Rev. Lett.* **85** 4988
[5] SantaLucia J Jr 1998 *Proc. Natl Acad. Sci. USA* **95** 1460 and references there in
[6] Yeramian E *et al* 1991 *Biopolymers* **30** 481 and references therein
[7] Ivanov V, Piontkovski D and Zocchi G 2005 *Preprint* cond-mat/0506067v1
[8] Everaers R *et al* 2007 *Phys. Rev.* E **75** 041918
[9] Zeng Y, Montrichok A and Zocchi G 2004 *J. Mol. Biol.* **339** 67–75
[10] Zeng Y and Zocchi G 2006 *Biophys. J.* **90** 4522–29
[11] Montrichok A, Gruner G and Zocchi G 2003 *Europhys. Lett.* **62** 452
[12] Zeng Y, Montrichok A and Zocchi G 2003 *Phys. Rev. Lett.* **91** 148101
[13] Ares S *et al* 2005 *Phys. Rev. Lett.* **94** 035504

[14] van Erp T S, Cuesta-Lopez S and Peyrard M 2006 *Eur. Phys. J.* E **20** 421
[15] Cule D and Hwa T 1997 *Phys. Rev. Lett.* **79** 2375–8
[16] Ivanov V, Zeng Y and Zocchi G 2004 *Phys. Rev.* E **70** 051907
[17] Ivanov V, Piontkovski D and Zocchi G 2005 *Phys. Rev.* E **71** 041909