Single-molecule derivation of salt dependent base-pair free energies in DNA

Josep M. Huguet[a], Cristiano V. Bizarro[a,b], Núria Forns[a,b], Steven B. Smith[c], Carlos Bustamante[c,d], and Felix Ritort[a,b,1]

[a]Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain; [b]CIBER-BBN de Bioingenieria, Biomateriales y Nanomedicina, Instituto de Sanidad Carlos III, Madrid, Spain; [c]Department of Physics, and [d]Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, CA 94720

Edited by Olke C. Uhlenbeck, Northwestern University, Evanston, IL, and approved July 19, 2010 (received for review February 4, 2010)

Accurate knowledge of the thermodynamic properties of nucleic acids is crucial to predicting their structure and stability. To date most measurements of base-pair free energies in DNA are obtained in thermal denaturation experiments, which depend on several assumptions. Here we report measurements of the DNA base-pair free energies based on a simplified system, the mechanical unzipping of single DNA molecules. By combining experimental data with a physical model and an optimization algorithm for analysis, we measure the 10 unique nearest-neighbor base-pair free energies with 0.1 kcal mol^{-1} precision over two orders of magnitude of monovalent salt concentration. We find an improved set of standard energy values compared with Unified Oligonucleotide energies and a unique set of 10 base-pair-specific salt-correction values. The latter are found to be strongest for AA/TT and weakest for CC/GG. Our unique energy values and salt corrections improve predictions of DNA unzipping forces and are fully compatible with melting temperatures for oligos. The method should make it possible to obtain free energies, enthalpies, and entropies in conditions not accessible by bulk methodologies.

DNA thermodynamics | DNA unzipping | nearest-neighbor model | optical tweezers

The nearest-neighbor (NN) model (1–4) for DNA thermodynamics has been successfully applied to predict the free energy of formation of secondary structures in nucleic acids. The model estimates the free-energy change to form a double helix from independent strands as a sum over all of resulting bp and adjacent-bp stacks, depending on the constituent four bases of the stack, by using 10 nearest-neighbor base-pair (NNBP) energies. These energies themselves contain contributions from stacking, hydrogen-bonding, and electrostatic interactions as well as configurational entropy loss. Accurately predicting free energies has many applications in biological science: to predict self-assembled structures in DNA origami (5, 6); achievement of high selectivity in the hybridization of synthetic DNAs (7); antigene targeting and siRNA design (8); characterization of translocating motion of enzymes that mechanically disrupt nucleic acids (9); prediction of nonnative states (e.g., RNA misfolding) (10); and DNA guided crystallization of colloids (11).

Some of the most reliable estimates of the NNBP energies to date have been obtained from thermal denaturation studies of DNA oligos and polymers (2). Although early studies showed large discrepancies in the NNBP values, nowadays they are remarkably consistent among several groups. In these studies it is assumed that duplexes melt in a two-state fashion. However this assumption is not often the case and a discrepancy between the values obtained using oligomers vs. polymers remains a persistent problem that has been attributed to many factors such as the slow dissociation kinetics induced by a population of transient nondenatured intermediates that develop during thermal denaturation experiments (12). Single-molecule techniques (13) circumvent such problems by allowing one to control and monitor the denaturated state of a molecule along a full reaction coordinate. This paper reports measurements of the 10 NNBP energies in DNA by mechanically melting individual DNA molecules using an advanced optical tweezers apparatus. By measuring the force-distance curves (FDCs) we can determine the free energy of formation for the double helix. Previous studies have suggested using single-molecule force measurements to extract the NNBP energies in a wide variety of conditions (14). Here we show how, by combining developments in optical tweezers technology with refined data analysis, it is possible to determine free-energy parameters with high precision (0.1 kcal/mol) in a wide variety of conditions, including salt concentration, pH, and temperature. In particular, we have derived the salt corrections that apply to a wider range of salt (0.01–1 M NaCl) compared to the ranges (0.05–1 M NaCl) explored in thermal denaturation experiments (2).

Results

Mechanical melting of DNA consists of pulling apart (unzipping) the two strands of a double-stranded DNA (dsDNA) molecule until the base pairs that hold the duplex together are disrupted and two single-stranded DNA (ssDNA) molecules are obtained. Such experiments reveal a FDC with a characteristic sawtooth pattern with force rips that are correlated with the DNA sequence (15–19). Features of this pattern are too coarse to distinguish individual base-pairs, but the energy of a particular type of bp stack can be inferred by its effect at many locations along the curve. To extract the NNBP energies high quality signal-to-noise measurements and reversible pulls are required. In previous studies of DNA unzipping either the length of the handles was too long (15, 18, 19) (permitting large thermal fluctuations in the molecular extension) or the experiments were performed at fast pulling speeds (16) and the unzipping/rezipping FDCs showed hysteresis (18) indicating that the process was not quasistatic. Here we demonstrate that by pulling DNA hairpins with extremely short handles at low pulling rates, one obtains FDCs that are essentially reversible (unzipping = rezipping). Besides, the slow pulling rate allows the system to visit states of higher energy at each fixed trap position. This fact permits one to obtain an estimation of the equilibrium FDC (see *Materials and Methods*). These experimental FDCs can then be compared quantitatively with synthetic FDCs generated *in silico* by a physical model. To perform these tasks we have developed a miniaturized dual-beam optical tweezers apparatus (20, 21) (see *SI Appendix: Sections S1, S2*) and a curve-alignment algorithm to cancel instrument drift. The physical model involves an algorithm for searching the 10-dimensional space of possible NNBP energies that gives rapid and robust convergence to an optimum fit.

Author contributions: J.M.H. and F.R. designed research; J.M.H., C.V.B., and N.F. performed research; J.M.H., C.V.B., S.B.S., and F.R. analyzed data; and J.M.H., S.B.S., C.B., and F.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: fritort@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001454107/-/DCSupplemental.

For the unzipping experiments, a molecular construct was synthesized starting from a 6,838 bp long DNA hairpin that was flanked by very short handles (29 bp) and a tetraloop (5′-ACTA-3′) at its end (Fig. 1*A* and *SI Appendix: Section S3*). The molecule is tethered between an optically trapped bead and a bead at the tip of a pipette held by suction (Fig. 1*B*). DNA molecules are unzipped by moving the optical trap at low pulling speed (10 nm/s) and the reversible FDC is measured (Fig. 1*C*). Fig. 1*D* shows the unzipping FDCs of one molecule at various salt concentrations. The equilibrium FDC ($F_{eq}(x_{tot})$) describing the experiments can be obtained by computing the partition function of the system, Z, at a total distance x_{tot} (see *Materials and Methods*),

$$Z(x_{tot}) = \sum_n \exp\left(-\frac{G(x_{tot},n)}{k_B T}\right);$$

$$F_{eq}(x_{tot}) = -k_B T \frac{\partial}{\partial x_{tot}} \ln Z(x_{tot}), \qquad [1]$$

where the sum in Z extends over all possible intermediate states (n). The free energy ($G(x_{tot},n)$) has three main contributions (see *SI Appendix: Fig.S1*). The first one is due to the stacking and hydrogen-bond energies of the bases, while the second one comes from the elastic contribution of the released ssDNA during unzipping. The third one is the parabolic potential that allows us to progressively unzip the molecular construct and access any particular region of the sequence. When dsDNA is melted mechanically, the two product strands of the reaction are produced under a tension around 15 pN, whereas a standard melting reaction produces ssDNA strands under zero tension. Therefore to achieve the same final state as a standard reaction, one needs to relax the separate strands and measure the energy that is returned. Strands of ssDNA exhibit a complex form of entropic elasticity which can be modeled numerically (22). A set of empirical fits to ssDNA elasticity was developed based on pulling experiments done with a 3 kb piece of ssDNA (Fig. 1*E* and *SI Appendix: Section S4*). We have found that the worm-like-chain (WLC) model correctly fits the ssDNA elastic response for salt concentrations below 100 mM NaCl, whereas above this value FDCs develop a plateau at low forces related to the formation of secondary structures (23). Above the force plateau the freely-jointed chain (FJC) model fits data better (*SI Appendix: Section S5*). In order to extract the NNBP free-energy changes for a molecule that adopts a (hypothetical) non-self-interacting melted state it is best to exclude the formation of partial secondary structures in the initial and final states. This exclusion can be achieved by interpolating a FJC fit between the high force region (e.g., above 15 pN) and the point of zero force, i.e., zero molecular extension. Here we are assuming that the ssDNA has no (secondary) structure at high salt concentrations. Our measurements give the free-energy difference between two ideal ssDNA complementary strands and the hybridized dsDNA duplex. This assumption holds at low salt concentrations and is an approximation at high salts. Discounting secondary structures in the unzipped "coil" state is an improvement over previous bulk methods where they could not detect such structures and these structures violate the two-state hypothesis (12). We have checked that the obtained elastic properties of the ssDNA match the last part of the unzipping FDC, when the molecule is fully extended (Fig. 2 *A* and *B*, magenta curve).

When we include the effects of elasticity and calculate the FDC by using the NNBP energies provided by SantaLucia (2) (i.e., the Unified Oligonucleotide (UO) energies) and currently used by Mfold (24–26), we observe a qualitative agreement with the experimental FDC (Fig. 2 *A* and *B*, black and blue curves). Nevertheless, slight but systematic deviations between both curves are observed, particularly when considering the effect of salt concentration. Although the mean unzipping force predicted by the UO energies shows a logarithmic dependence with salt concentration (Fig. 2*C*, green curve) it overestimates the measured values by nearly 1 pN at low salt. These differences indicate a slight error in the NNPB UO energies. According to Eq. **1** an average 8% correction in the NNBP energies introduces a difference of 1 pN in the mean unzipping force. The best values for the NNBP energies ($\varepsilon_i, i = 1,...,10$) can be inferred by minimizing the mean squared error between the experimental and theoretical FDCs:

$$E(\varepsilon_1,\ldots,\varepsilon_{10},\varepsilon_{loop}) = \frac{1}{N}\sum_{i=0}^{N}(F_i^{exp} - F_i^{the}(\varepsilon_1,\ldots,\varepsilon_{10},\varepsilon_{loop}))^2, \qquad [2]$$

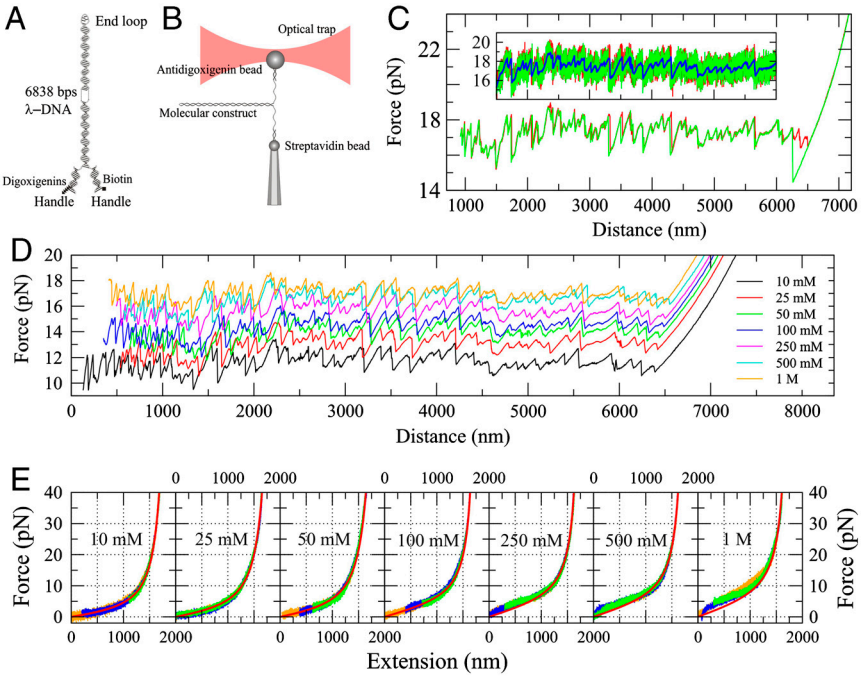

Fig. 1. Experimental setup and results. (*A*) Molecular construct. A sequence of 6,838 bp obtained from λ-DNA is ligated to a tetraloop and two short handles of 29 bp each. (*B*) Experimental setup. The molecular construct is attached to two beads. The unzipping experiment is performed by moving the optical trap relative to the pipette. (*C*) Unfolding (red) and refolding (green) curves filtered with a running average filter of 1 Hz bandwidth. There is always some hysteresis in the last rip. Pulling and relaxing are almost identical in the rest of the curve. Inset shows raw data (same color as before, blue curve is the average at bandwidth 1 Hz). (*D*) FDCs at various monovalent salt concentrations. (*E*) Elastic response of a 3 kb ssDNA molecule at various salts. Raw data of three molecules are shown (orange, green and blue curves). Red curve shows the best-fit to the elastic model.

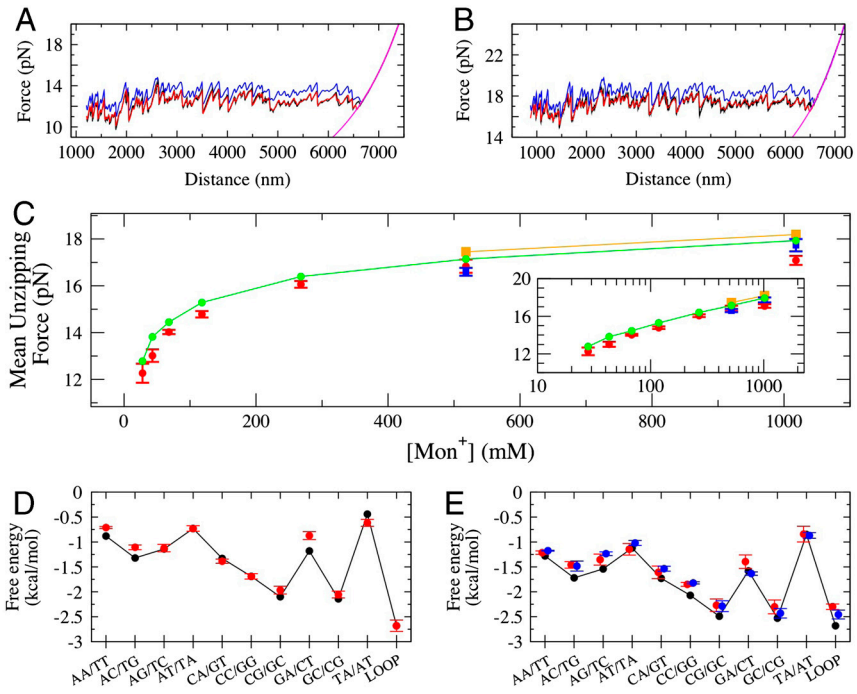

Fig. 2. Salt dependencies. (*A*, *B*) FDCs for the 6.8 kb sequence at 10 mM NaCl (*A*) and 1 M NaCl (*B*). Black curve, experimental measurements; blue curve, UO prediction; red curve, our fit; magenta curve, elastic response of the fully unzipped molecule. The theoretical FDC is calculated in equilibrium, which assumes that the bandwidth is 0 Hz and the experimental data is filtered at bandwidth 1 Hz. If data is filtered at higher frequencies (>1 Hz), hopping between states is observed and the experimental FDC does not compare well with the theoretical FDC at equilibrium. If data is filtered at lower frequencies (<1 Hz), the force rips are smoothed and hopping transitions are averaged out. (*C*) Dependence of mean unzipping force with salt concentration. Red points, experimental measurements for the 6.8 kb sequence; green curve, UO prediction for the 6.8 kb sequence; blue points, experimental measurements for the 2.2 kb sequence; orange curve, UO prediction for the 2.2 kb sequence. (*D*, *E*) NNBP energies and comparison with UO values at 10 mM NaCl (*D*) and 1 M NaCl (*E*). The following notation is used for NNBP: AG/TC denotes 5′-AG-3′ paired with 5′-CT-3′. Black points, UO values; red points, values for the 6.8 kb molecule; blue points, values for the 2.2 kb molecule. The values for the 6.8 kb and the 2.2 kb molecules have been obtained after averaging over six molecules. Error bars are determined from the standard error among different molecules.

where $E(\varepsilon_1,\ldots,\varepsilon_{10},\varepsilon_{\text{loop}})$ is the total error, N is the number of experimental points in the FDC, and F_i^{exp} is the experimental equilibrium force, averaged with a low-pass filter of 1 Hz to avoid hopping artifacts. In order to extract the best values for the ε_i, Eq. **2** is minimized in an 11-dimensional space using a Monte Carlo based approach (see *SI Appendix: Section S6*). Moreover, in order to correct for the position drift in the optical tweezers, a shift function in the model was introduced that locally shifts the position of the trap along the FDC. The shift function leads to an improved match between theoretical and experimental FDCs without affecting the NNBP energy values (*SI Appendix: Section S7*). Our best-fit energy parameters reduce the error (i.e., give improved agreement) between the measured FDC and the theoretical prediction (Fig. 2 *A* and *B*, red and black curves). By minimizing the error function Eq. **2** and estimating the NNBP energies for many individual molecules, we can obtain error limits on the free-energy values. In Fig. 2 *D* and *E* we plot the average value and the standard error of the NNBP obtained for the 6.8 kb sequence and the UO prediction for the NNBP energies at 10 mM and 1 M NaCl. It is interesting that some of the new values are in good agreement with the results computed by SantaLucia (2) (e.g., CA/GT and AT/TA motifs) while others differ significantly (e.g., AA/TT and GA/CT at 10 mM NaCl and AC/TG and CC/GG at 1 M NaCl). According to the UO salt correction, the NNBP energies are extrapolated homogenously (i.e., the same salt correction is taken for all base-stack combinations) from standard salt conditions (1 M NaCl) down to lower salt concentrations (e.g., 50 mM). However, such correction does not predict the observed unzipping force at low salt, especially for certain NNBP such as AA/TT or GA/CT. This discrepancy is somehow expected at the lowest salt regime (10–50 mM NaCl) since the UO salt correction applies between 0.05–1 M NaCl. Nevertheless, we also observe discrepancies in the mean unzipping force at salt concentrations above 100 mM NaCl. A heterogeneous (sequence specific) salt correction could provide consistent results with the experiment. Such deviations are not unexpected, given the differences in solvation between specific nucleotides and salt ions (27, 28), however the effect has never been quantified in the context of polymeric DNA. With this goal in mind, we applied our fitting algorithm to extract NNBP energies for data taken at many salt concentrations (Fig. 3, red points). As a further check, we repeated the experiments at 500 mM and 1 M NaCl by using a different test sequence from the other end of lambda (see *SI Appendix: Section S3*). Thus we found compatible NNBP energies between the two test molecules (Fig. 2*E* and Fig. 3, blue dots).

The UO model uses a nonsequence-specific salt concentration correction for the different NNBP energies given by:

$$\varepsilon_i([\text{Mon}^+]) = \varepsilon_i^0 - m \cdot \ln([\text{Mon}^+]), \qquad [3]$$

where $\varepsilon_i([\text{Mon}^+])$ is the energy of formation of the ith NNBP ($i = 1,\ldots,10$) at monovalent salt concentration $[\text{Mon}^+]$ (expressed in M units), $\varepsilon^0{}_i$ is the NNBP energy at 298 K, 1 M monovalent salt and m is the nonspecific prefactor equal to $m = 0.110$ kcal/mol (2, 25) at 298 K (Fig. 3, green lines). The

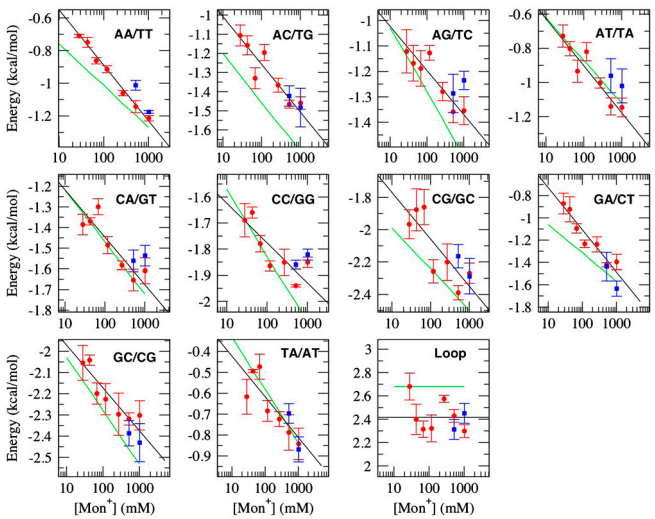

Fig. 3. Salt corrections of the NNBP energies. Figure shows the energy of all different NNBP parameters. Red (blue) points are the experimental results for the 6.8 kb (2.2 kb) sequence; green curve, UO nonspecific salt correction; black curve, fit to Eq. **3** with adjustable parameters m_i ($i = 1,\ldots,10$, *loop*) and $\varepsilon^0{}_i$.

UO model assumes that only the entropy (and not the enthalpy) depends on the salt and this dependence is uniform at all temperatures meaning that $m(T)/T$ is a constant. Therefore, the correction of salt for the free-energy changes depends on the temperature according to $m(T) = \frac{T}{T_0} m(T_0)$ where T_0 is a reference temperature (see *SI Appendix: Section S8*). To make a heterogeneous salt correction, one needs only to define 10 sequence-specific prefactors m_i to be used with the same logarithmic dependence as shown in Eq. **3**. Thus we fit all NNBP energies using NNBP-dependent parameters m_i ($i = 1,\ldots,10$, *loop*) and ε^0_i. Such a fit is shown in Fig. 3 (black lines) and best values for m_i are listed in Table 1. We observe that the salt dependence of some NNBP parameters is well described by the UO nonspecific correction (e.g., AT/TA and CA/GT) but most of them are better fit with some correction in parameters ${\varepsilon^0}_i$ and m (e.g., AA/TT, AC/TG, AG/TC). We have noticed that NNBP purine-purine or pyrimidine-pyrimidine combinations (5′-YY-3′ or 5′-RR-3′, i.e., AA/TT, AG/TC, CC/GG, GA/CT) differ most from the UO homogeneous salt correction than mixed purine-pyrimidine combinations (5′-RY-3′ and 5′-YR-3′). A difference between these combinations can be observed in how charges (e.g., hydrogen-bond acceptor and donor groups) are distributed along the major group of the double helix. The latter have charged groups that tend to be uniformly distributed between the two strands along the major groove, whereas the former have donor and acceptor groups unevenly distributed between the two strands. The specific salt correction found in our measurements could be consequence of how monovalent cations bind the two strands along the major groove. There are precedents to such results: Sugimoto and coworkers (29) have reported that cation binding is correlated to duplex stability. Computer simulations have identified acceptor groups in guanine (N7, O6) and adenine (N7) as preferential cation binding sites (30). Our experimentally determined specific salt corrections might be interpreted as a corroboration of such hypothesis.

Finally, we wished to check how well our unique free-energy values work to predict the melting temperature of oligonucleotides under various salt conditions. Fortunately, there are several published studies giving accurate experimental values. Although T_M is not a robust estimator to compare the melting and unzipping experiments, this is the most reliable experimental observable from melting data with which we can compare our results. For nonself-complementary oligos, the melting temperatures can be estimated from the following expression (2)

$$T_M = \frac{\Delta H^0}{\Delta S^0 + \sum_i \frac{m_i(T)}{T} \ln[\mathrm{Mon}^+] + R \ln[C_T/4]}, \qquad [4]$$

where ΔH^0 and ΔS^0 are the oligo enthalpy and entropy that are assumed to be temperature independent, m_i are taken at $T = 298$ K and C_T is the total single-stranded concentration of the oligo. In order to compare UO free energies with the unique values, we have recalculated the melting temperatures of 92 oligos at five different salt conditions and compared our results with melting data taken from Owczarzy et al. (31). By taking the UO values for the NNBP free energies, enthalpies, and entropies (with the corresponding initiation factor for each oligo, see *SI Appendix: Section S8 and Table S1*) at standard conditions (1 M NaCl) but using the heterogeneous salt correction, the error committed in the extrapolation at lower salts is found to be below 2 °C, which is similar to the error of the UO model (Fig. 4*A*). This fact reveals that the average salt correction for all NNBP, $m = \frac{1}{16}\sum_{i=0}^{16} m_i = 0.104$, is equivalent to the homogeneous UO correction ($m = 0.110$) at 25 °C. However, a closer inspection shows that a heterogeneous salt correction does a better job in predicting melting temperatures than the homogeneous one (Fig. 4*B*) for oligos longer than 15 bp. We have fixed the 10 values for the NNBP free energies ε_i and the 10 parameters m_i as given by our measurements and determined the 10 enthalpies Δh_i that minimize the error function $\chi^2 = \frac{1}{N}\sum_i^N (T_i^{\mathrm{exp}} - T_i^{\mathrm{pred}})^2$, where i runs over all ($N = 460$) oligos and salt conditions shown in ref. 31 (see *SI Appendix: Section S9*). Here T_i^{exp} is the melting temperature experimentally measured in ref. 31 and T_i^{pred} is the melting temperature predicted by Eq. **4**. For oligos longer than 15 bp, the UO parameters with homogeneous salt corrections give $\chi^2 = 2.37$ (corresponding to 1.5 °C average error) whereas the optimal enthalpies (Table 2) give a lower error, $\chi^2 = 1.74$ (1.3 °C average error). For oligos of 15 bp or shorter our best values underestimate melting temperatures by 2–4 °C. Finally we note that a moderate increase of the NNBP values by 0.15 kcal/mol (i.e., slightly beyond the standard error given for the NNBP values) makes the standard deviation error for temperature melting prediction go from ≅2 °C up to 5–6 °C.

Discussion

Why do our free-energy numbers predict fairly well melting temperatures of oligos longer than 15 bp but do worse for shorter ones? Discrepancies between predicted and measured melting temperature for short oligos have been already reported in bulk measurements (32) and attributed to differences in analytical methods used to extract melting temperatures. Another possible explanation is that short oligos (≤15 bp) might not have the double helix perfectly formed and the formation energies involved in the duplex are slightly different from the energies for longer sequences. Although we lack a conclusive answer to this question, it is worth underlining that UO free-energy values are obtained in order to correctly predict the melting temperatures for all oligo lengths. This constraint might lead to error compensation between the melting temperature datasets corresponding to short and long oligos. Let us stress that with increasing length, the

Table 1. Summary of results at 298 K, 1 M [NaCl]

NNBP	6 kb	2 kb	Best	UO	${\varepsilon^0}_i$	m_i
AA/TT	−1.21 (0.02)	−1.18 (0.01)	−1.20 (0.02)	−1.27	−1.23 (0.01)	0.145 (0.006)
AC/TG	−1.46 (0.04)	−1.48 (0.10)	**−1.47 (0.10)**	−1.71	−1.49 (0.05)	0.10 (0.02)
AG/TC	−1.35 (0.07)	−1.24 (0.04)	**−1.30 (0.07)**	−1.53	−1.36 (0.03)	0.070 (0.014)
AT/TA	−1.15 (0.06)	−1.02 (0.05)	−1.09 (0.08)	−1.12	−1.17 (0.04)	0.12 (0.02)
CA/GT	−1.61 (0.07)	−1.54 (0.05)	−1.58 (0.08)	−1.72	−1.66 (0.05)	0.09 (0.02)
CC/GG	−1.85 (0.02)	−1.82 (0.02)	**−1.84 (0.03)**	−2.08	−1.93 (0.04)	0.06 (0.02)
CG/GC	−2.27 (0.06)	−2.29 (0.10)	**−2.28 (0.12)**	−2.50	−2.37 (0.09)	0.13 (0.04)
GA/CT	−1.40 (0.07)	−1.63 (0.04)	−1.50 (0.08)	−1.57	−1.47 (0.05)	0.15 (0.02)
GC/CG	−2.30 (0.06)	−2.43 (0.10)	−2.36 (0.11)	−2.53	−2.36 (0.04)	0.08 (0.02)
TA/AT	−0.84 (0.08)	−0.87 (0.06)	−0.85 (0.10)	−0.84	−0.84 (0.05)	0.09 (0.02)
Loop	2.30 (0.06)	2.46 (0.09)	2.37 (0.10)	2.68	2.43 (0.05)	-

Free energies are given in kcal/mol. 6 kb (2 kb) are the energies obtained from the averaged results from the 6.8 kb (2.2 kb) sequences (standard error in parenthesis). Best is an average of the 2.2 kb and 6.8 kb results. In bold type letter we highlight the bps that disagree most with the values predicted by the UO model (extracted from ref. 2). ${\varepsilon^0}_i$ and m_i are the standard energies and prefactors obtained from the fits of Eq. **3**, shown in Fig. 3

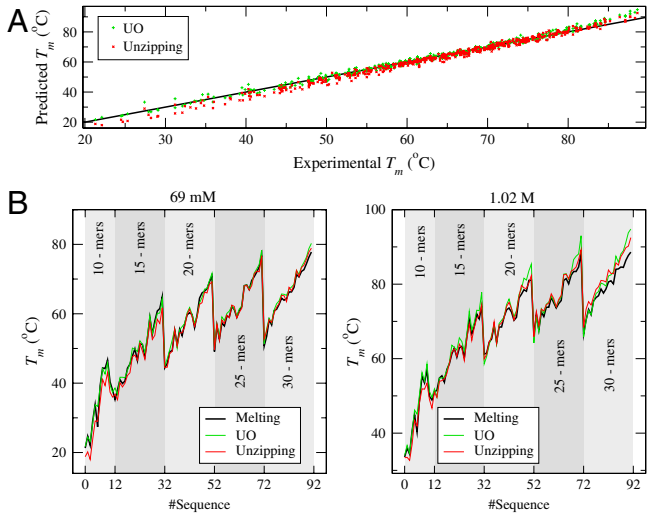

Fig. 4. Melting temperatures prediction. Comparison with melting temperatures for the 92 oligos ranging from 10–30 bp reported in ref. 31 (data reported in *SI Appendix: Table S2*). (*A*) Predicted vs. experimentally measured melting temperatures at five salt conditions ([Na^+] = 69, 119, 220, 621, and 1,020 mM). The values obtained from unzipping have less error at higher temperatures (corresponding to longer oligos). (*B*) Prediction at 69 mM (left) and 1.02 M NaCl (right). Black lines are the experimentally measured melting temperatures, green line is the UO prediction and red line our prediction from unzipping data.

T_M prediction is more tolerant of errors in the details of the NNBP energies, where sequence effects are averaged out. In addition, deviations from the bimolecular model (*SI Appendix: Section S8, Eq. S9*) arise for sequences with $n \geq 20$, as their melting process begins to shift toward pseudomonomolecular behavior (33). Still, our predicted melting temperatures for oligos with $n \geq 20$ agree well for the sequences reported in ref. 31.

We have performed single-molecule force unzipping experiments to extract DNA bp free energies at various salt concentrations finding heterogeneous salt corrections. What is the origin of specific salt corrections? As previously said, this specificity might be consequence of how donor and acceptor groups distribute between the two strands along the major groove of the helix. However, there is an alternative interpretation based on the sequence dependence of ssDNA elasticity. Previous studies (34) have suggested a conformational transition of the sugar pucker in ssDNA that goes from the A-form (C3′-endo) at low forces to the B-form (C2′-endo) at high forces. A related phenomenon has been reported in recent stretching studies of homopolymeric RNA sequences (35) that reveal sequence dependent base-stacking effects. Based on our experimental data we cannot discard such interpretation. An exhaustive research of the elastic response of different homopolymeric ssDNA sequences (spanning different combinations of stacked bases) could shed light into this question.

The unique values for the NNBP energies reported here are compatible with force unzipping experiments and improve melting temperature prediction as compared to the UO prediction for oligos longer than 15 bp. Although melting and unzipping experiments are based on disruption processes triggered by different external agents (temperature and force respectively), the agreement between the NNBP energies obtained is remarkable. Our work shows that using very different experimental systems the NN model can provide remarkably consistent results. The unique NNBP energies predict both the unzipping experiments and the melting temperatures of oligos fairly well at low temperatures and low salt concentration. In these experimental conditions, the unzipping experiments provide an alternative determination of the NNBP parameters that seems to work better than the UO parameters. One important advantage over optical melting experiments is that the folding/unfolding transition does not need to be two-state. Besides, instead of several short oligos of different sequence, one long molecule is sufficient to infer the NNBP energies. The main limitation of our method is the accurate determination of the elastic response of the ssDNA. A 10% error in the estimation of the persistence or Kuhn length of ssDNA induces a similar error in the prediction of the NNBP energies. Moreover, the bimolecular initiation factors cannot be determined with our methodology. This approach can be extended to extract free energies, entropies, and enthalpies in DNA and RNA structures under different solvent and salt conditions. To estimate NNBP entropies and enthalpies we should have to perform experiments at different temperatures. At present this cannot be achieved with our experimental setup because the changes of temperature dramatically affect the optics of the instrument. Temperature variations introduce undesirable drift effects that compromise the resolution of the measurements. The method can be also applied to extract free energies of other structural motifs in DNAs (e.g., sequence dependent loops, bulges, mismatches, and junctions). Most important, force methods make it possible to extract free energies in conditions not accessible to bulk methods. One example is unzipping dsRNA in the presence of magnesium where free-energy prediction is not possible from melting experiments because RNA hydrolyzes below the melting temperature. Another example is binding free energies of DNAs and RNAs bound to proteins where the proteins denaturalize below the dissociation melting transition. Although bind-

Table 2. Melting temperature prediction and optimal enthalpies and entropies.

Method	UO values				Force measurements			
NNBP	ε_i 25 °C	Δh_i	Δs_i	m_i	ε_i25 °C	Δh_i	Δs_i	m_i
AA/TT	−1.28	−7.9	−22.2	0.114	−1.23	**−7.28 (0.3)**	−20.28 (1.2)	0.145
AC/TG	−1.72	−8.4	−22.4	0.114	−1.49	**−5.80 (0.3)**	−14.46 (1.3)	0.099
AG/TC	−1.54	−7.8	−21.0	0.114	−1.36	**−5.21 (0.3)**	−12.89 (1.2)	0.070
AT/TA	−1.12	−7.2	−20.4	0.114	−1.17	**−4.63 (0.6)**	−11.62 (2.1)	0.117
CA/GT	−1.73	−8.5	−22.7	0.114	−1.66	**−8.96 (0.3)**	−24.48 (1.2)	0.091
CC/GG	−2.07	−8.0	−19.9	0.114	−1.93	**−8.57 (0.3)**	−22.30 (1.2)	0.063
CG/GC	−2.49	−10.6	−27.2	0.114	−2.37	**−9.66 (0.5)**	−24.43 (2.1)	0.132
GA/CT	−1.58	−8.2	−22.2	0.114	−1.47	**−8.16 (0.3)**	−22.46 (1.3)	0.155
GC/CG	−2.53	−9.8	−24.4	0.114	−2.36	**−10.10 (0.5)**	−25.96 (1.8)	0.079
TA/AT	−0.85	−7.2	−21.3	0.114	−0.84	**−8.31 (0.6)**	−25.06 (2.1)	0.091
χ^2		2.37				1.74		

Free energies and enthalpies given in kcal/mol; entropies given in cal/mol · K. UO values at standard conditions plus homogeneous salt corrections (left block of columns) have larger χ^2 error in predicting melting temperatures for oligos longer than 15 bp of ref. 31 than heterogeneous salt corrections (right block of columns). Combining melting temperature data from ref. 31 and the new values for $\varepsilon^0{}_i$ and m_i we can also extract the optimal enthalpies Δh_i (highlighted) and entropies Δs_i. The values in parenthesis indicate the range of Δh_i and Δs_i that also predict the melting temperatures of the oligos within an average error of 2 °C (i.e., the typical experimental error claimed in melting experiments).

ing/unbinding FDCs could possibly not be reversible, nonequilibrium free-energy methods could be used to extract the binding free energy. This method could be also useful in cases where molecular aggregation and other collective effects in bulk preclude accurate free-energy measurements. Finally, the NN model (with adjusted parameters) has predicted the experimental FDC data very well, but we have only explored two specific sequences from lambda. By using other dsDNA molecules, we could search for long-range context effects (e.g., second-nearest and third-nearest interactions) in precisely those places along the sequence where the present model might be seen to fail. Because the mechanical unzipping method can be performed in many sequences it may be time to test the applicability of the next-nearest-neighbor model.

Our work establishes a unique methodology to obtain thermodynamic information from single-molecule experiments. Future experiments will address the question of the sequence-specific salt effects in the ssDNA. A different expansion of our work will determine the NNBP enthalpies by performing experiments at different temperatures. The use of stiffer optical traps, which makes the partition function more localized, will provide more precise energy measurements. Stiffer optical traps will be a starting point to observe second NN effects.

Materials and Methods

The FDC is calculated by using a mesoscopic model that describes separately each component of the experimental setup (36): the bead in the optical trap, the handles, the released ssDNA, and the dsDNA hairpin (see *SI Appendix: Fig. S1*). The potential energy of the bead in the optical trap is described by a harmonic potential which is determined by the stiffness of the trap $E_b(x_b) = \frac{1}{2}kx_b^2$, where k is the trap stiffness and x_b is the elongation of the bead from the center of the trap. We use the NN model to describe the free energy of formation of the DNA duplex (2). Since DNA has four types of bases (adenine, guanine, cytosine, and thymine), the NN model should have 16 different parameters. However, due to symmetry considerations, there are only 10 independent parameters. The free energy required to open n bps is given by the sum of the free energies required to open each consecutive NN pair $G_{\mathrm{DNA}}(n) = \sum_{i=0}^{n} \varepsilon_i$, where $G_{\mathrm{DNA}}(n)$ is the free energy of the hairpin when n bps are disrupted and ε_i is the free energy required to disrupt the bp i. Therefore, the free energy of formation of the duplex depends on the sequence of bp. The NN model assumes that only local interactions between bps are relevant. It is also assumed that each interaction (composed of hydrogen bonding, stacking, and entropy loss) can be described by one single free-energy value. An extra free-energy contribution is included in the model to account for the disruption of the end loop (see *SI Appendix: Section. S10*). Elastic models for polymers are used to describe the elasticity of the handles and the ssDNA released during the unzipping process. The handles are dsDNA and they are modeled using the force vs. extension curve of a WLC $F(x_h) = \frac{k_B T}{4l_p}((1-\frac{x_h}{L_0})^{-2} - 1 + 4\frac{x_h}{L_0})$, where k_B is the Boltzmann constant and T is the temperature, l_p is the persistence length and L_0 is the contour length. The elastic free energy of the handles is obtained by integrating the previous expression. The ssDNA is modeled using either a WLC or a FJC model, depending on the salt concentration of the experiment (see *SI Appendix: Section S5*). In the case of the FJC model, the following equation gives the extension vs. force curve, $x_s(F) = L_0(\coth(\frac{bF}{k_B T}) - \frac{k_B T}{bF})$, where b is the Kuhn length. Again, the elastic free energy of the ssDNA is obtained by integrating the force vs. molecular extension curve. The parameters that define the elastic response of the handles are taken from the literature (34): $l_p = 50$ nm and $L_0 = 9.86$ nm (=0.34 nm/bp × 29 bp). The total free energy of the total system is given by the sum of all free-energy contributions $G(x_{\mathrm{tot}},n) = E_b(x_b) + 2G_h(x_h) + 2G_s(x_s,n) + G_{\mathrm{DNA}}(n)$. The total distance of the system is given by the sum of all extensions corresponding to the different elements $x_{tot} = x_b + 2x_h + 2x_s$ (see *SI Appendix: Fig. S1*). The total free energy of the system is completely determined by x_{tot} and n. The equilibrium FDC can be numerically calculated via the partition function defined by Eq. **1** and where we only include sequential configurations (*SI Appendix: Section S11*). *SI Appendix: Section S12* describes the thermodynamic process of unzipping.

ACKNOWLEDGMENTS. We thank T. Betz, M. Orozco, J. Subirana, and I. Tinoco Jr. for useful comments and a careful reading of the manuscript. J.M.H was supported by the Spanish Research Council in Spain. F.R. is supported by Grants FIS2007-3454 and Human Frontier Science Program (HFSP) (RGP55-2008).

1. Calladine C, Drew H, Luisi B, Travers A (2004) *Understanding DNA: the molecule and how it works* (Elsevier Academic Press, Amsterdam), 3rd Ed.
2. SantaLucia JJ (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* 95:1460–1465.
3. Devoe H, Tinoco IJ (1962) The stability of helical polynucleotides: base contributions. *J Mol Biol* 4:500–517.
4. Crothers DM, Zimm BH (1964) Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. *J Mol Biol* 9:1–9.
5. Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440:297–302.
6. Douglas SM, et al. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 459:414–418.
7. Shoemaker DD, et al. (2001) Experimental annotation of the human genome using microarray technology. *Nature* 409:922–927.
8. Dorsett Y, Tuschl T (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* 3:318–329.
9. Johnson DS, Bai1 L, Smith BY, Patel SS, Wang MD (2007) Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* 129:1299–1309.
10. Russel R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1–20.
11. Nykypanchuk D, Maye MM, van der Lelie D, Gang O (2008) DNA-guided crystallization of colloidal nanoparticles. *Nature* 451:549–552.
12. Holbrook JA, Capp MW, Saecker RM, Record MTJ (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* 38:8409–8422.
13. Ritort F (2006) Single-molecule experiments in biological physics: methods and applications. *J Phys Condens Mat* 18:R531–R583.
14. Lubensky DK, Nelson DR (2000) Single molecule statistics and the polynucleotide unzipping transition. *Phys Rev E* 65:031917.
15. Essevaz-Roulet B, Bockelmann U, Heslot F (1997) Mechanical separation of the complementary strands of DNA. *Proc Natl Acad Sci USA* 94:11935–11940.
16. Rief M, Clausen-Schaumann H, Gaub HE (1999) Sequence-dependent mechanics of single DNA molecules. *Nat Struct Biol* 6:346–349.
17. Danilowicz C, et al. (2003) DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc Natl Acad Sci USA* 100:1694–1699.
18. Bockelmann U, Thomen P, Essevaz-Roulet B, Viasnoff V, Heslot F (2002) Unzipping DNA with optical tweezers: high sequence sensitivity and force flips. *Biophys J* 82:1537–1553.
19. Shundrovsky A, Smith CL, Lis TJ, Peterson CL, Wang M (2006) Probing SWI/SNF remodeling of the nucleosome by unzipping single DNA molecules. *Nat Struct Biol* 13:549–554.
20. Bustamante C, Smith SB (2006) Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optical axis. *U.S. Patent* 7,133,132 B2.
21. Smith SB, Rivetti C (2010) TweezersLAB. http://tweezerslab.unipr.it.
22. Zhang Y, Zhou H, Ou-Yang Z-C (2001) Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. *Biophys J* 81:1133–1143.
23. Dessinges M, et al. (2002) Stretching single stranded DNA, a model polyelectrolyte. *Phys Rev Lett* 89:248102.
24. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415.
25. Peyret N (2000) Prediction of nucleic acid hybridization: parameters and algorithms. PhD dissertation (Wayne State University, Department of Chemistry, Detroit, MI).
26. Zuker M (2009) The Rensselaer bioinformatics web server. http://mfold.bioinfo.rpi.edu.
27. Tan Z-J, Chen S-J (2006) Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length. *Biophys J* 90:1175–1190.
28. Alemayehu S, et al. (2009) Influence of buffer species on the thermodynamics of short DNA duplex melting: sodium phosphate versus sodium cacodylate. *J Phys Chem B* 113:2578–2586.
29. Nakano S, Fujimoto M, Hara H, Sugimoto N (1999) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res* 27:2957–2965.
30. Cheatham TE, III, Kollman PA (1997) MD simulations highlight the structural differences among DNA:DNA,RNA:RNA and DNA:RNA hybrid duplexes. *J Am Chem Soc* 119:4805–4825.
31. Owczarzy R, et al. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry* 43:3537–3554.
32. Owczarzy R (2005) Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys Chem* 117:207–215.
33. Marky LA, Breslauer KJ (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26:1601–1620.
34. Smith SB, Cui Y, Bustamante C (1996) Overstretching B. DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271:795–799.
35. Seol Y, Skinner GM, Visscher K (2007) Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* 98:158103.
36. Manosas M, Ritort F (2005) Thermodynamic and kinetic aspects of RNA pulling experiments. *Biophys J* 88:3224–3242.