Published online 5 February 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 7* ***2405–2410***
doi:10.1093/nar/gkp016

A nonlinear dynamic model of DNA with a sequence-dependent stacking term

Boian S. Alexandrov[1], Vladimir Gelev[2], Yevgeniya Monisova[2], Ludmil B. Alexandrov[2], Alan R. Bishop[1], Kim Ø. Rasmussen[1,*] and Anny Usheva[2,*]

[1]Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545 and [2]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

Received November 18, 2008; Revised January 6, 2009; Accepted January 7, 2009

ABSTRACT

No simple model exists that accurately describes the melting behavior and breathing dynamics of double-stranded DNA as a function of nucleotide sequence. This is especially true for homogenous and periodic DNA sequences, which exhibit large deviations in melting temperature from predictions made by additive thermodynamic contributions. Currently, no method exists for analysis of the DNA breathing dynamics of repeats and of highly G/C- or A/T-rich regions, even though such sequences are widespread in vertebrate genomes. Here, we extend the nonlinear Peyrard–Bishop–Dauxois (PBD) model of DNA to include a sequence-dependent stacking term, resulting in a model that can accurately describe the melting behavior of homogenous and periodic sequences. We collect melting data for several DNA oligos, and apply Monte Carlo simulations to establish force constants for the 10 dinucleotide steps (CG, CA, GC, AT, AG, AA, AC, TA, GG, TC). The experiments and numerical simulations confirm that the GG/CC dinucleotide stacking is remarkably unstable, compared with the stacking in GC/CG and CG/GC dinucleotide steps. The extended PBD model will facilitate thermodynamic and dynamic simulations of important genomic regions such as CpG islands and disease-related repeats.

INTRODUCTION

The stability of the DNA double helical structure is determined by the interplay of a host of interactions, including hydrogen bonding, aromatic base stacking, backbone conformational constraints and electrostatic interactions, and the coordination of water molecules and metal ions. These interactions depend on DNA structure in a strongly nonlinear fashion. Not surprisingly, no simple model exists that accurately describes the melting behavior of double stranded (ds) DNA as a function of nucleotide sequence. A relatively successful strain of models for dsDNA denaturation originated from McClare's (1) ideas about the importance of nonlinear dynamics of vibrational excited states in macromolecules, followed by Davydov's prediction that vibrational excitations can be localized and trapped in a long lived state now known as the Davydov soliton (2). Inspired by these ideas Englander *et al*. (3) proposed that nonlinearity-induced localization of vibrational energy in dsDNA may lead to local DNA melting (DNA bubbles). This work in turn lead to, the Peyrard–Bishop–Dauxois (PBD) model of DNA (4,5), which uses a simple sequence-dependent nonlinear Hamiltonian to represent the separation of the dsDNA strands during a melting transition. The PBD model is unique among DNA models in that it combines elements from efficient *ad hoc* models of DNA melting and physically meaningful fully atomistic molecular dynamic simulations. Calculations based on the PBD model have successfully reproduced DNA melting and unzipping data for a variety of DNA sequences (6–12) but the model has also been used to calculate the probability for transient melting of the adenovirus genome (13) and to simulate the breathing dynamics of 100 bp long eukaryotic promoter regions (B. S. Alexandrov *et al*., submitted for publication). The sequence-dependence of the PBD Hamiltonian is encoded by an anharmonic term, which represents the H-bonding of a G–C or an A–T base pair. A nonlinear term representing the stacking energy of the consecutive base pairs mimics long-range effects along the DNA helix, but uses a sequence-independent average stacking force constant. The sequence-independence of the stacking term reflects the observation that the ratio of G–C to A–T hydrogen bonded base pairs is

*To whom correspondence should be addressed. Tel: +1 617 632 0522; Fax: +1 617 632 2927; Email: ausheva@bidmc.harvard.edu
Correspondence may also be addressed to Kim Ø. Rasmussen. Tel: +1 505 665 3851; Fax: +1 505 665 4063; Email: kor@lanl.gov

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

the strongest predictor of dsDNA stability. However, it is well known that relatively subtle differences in the stacking energies of the 10 dinucleotides steps can amount to very significant deviations in the melting temperature expected from a given G/C content (14–17). The best known examples of this phenomenon are the large differences in the melting temperatures of poly(dA).(dT) versus poly(dAdT), and poly(dG).(dC) versus poly(dGdC) (18).

Existing thermodynamic models of DNA melting that incorporate a nearest neighbor (NN) stacking energy term (16,19,20) perform significantly better than simple regression fit empirical formulas for DNA melting temperature. However, homogenous and periodic DNA sequences exhibit cumulative deviations from the ideal B-helix dsDNA structure, which result in further deviations in melting behavior that are difficult to account for by additive thermodynamic contibutions. Dynamic models of DNA, which represent long-range effects offer an advantage in this respect. In addition, such models offer certain mechanistic insight into the initial dynamically governed stages of DNA melting (DNA breathing), that have been implicated as relevant to protein binding and transcription initiation (B. S. Alexandrov *et al.*, submitted for publication; 7,13,21,22). The large deviations in the melting behavior of repeats and homopolymers was first reported in 1970 (18) and has since been discussed at length in the literature due to the abundance of such sequences in vertebrate genomes (23). However, potentially baring computationally intensive fully atomistic simulations, no method exists for analysis of the DNA breathing dynamics of repeats and highly G/C- or A/T-rich regions. Here we extend the original PBD model to include a sequence-dependent stacking term, in order to study such sequences. We collect melting data for several homogenous and periodic DNA oligos, and apply Monte Carlo (MC) simulations to derive 10 (24) distinct stacking force constants for the PBD Hamiltonian. The resulting PBD model has the potential to be applied at the genomic scale to conduct thermodynamic and dynamic simulations of important genomic regions such as CpG islands and disease related repeats.

MATERIALS AND METHODS

DNA oligonucleotides

The sequences of the 36 bp oligonucleotides used in the melting studies are listed in Table 1. The sequence of the L60B36 used in the MC simulations is 5′-CCGCCAGCG GCGTTATTACATTTAATTCTTAAGTATTATAAGT AATATGGCCGCTGCGCC-3′. The sequence of the P5 promoter used in the Langevin simulations is 5′-ACGCT GGGTATTTAAGCCCGAGTGAGCACGCAGGGTC TCCATTTTGAAGCGGGAGGTTTGAACGCGCAG CC-3′. The underlined AT residues at the transcriptional start site are replaced with GC in the transcriptionally silent P5 mutant promoter (7).

Table 1. Experimentally determined melting temperatures of dsDNA oligos that are used in the simulations

dsDNA	Melting Temperature, T_m (°C)	Stacking constant
$(G)_{36}.(C)_{36}$	74	k_{GG}
$(GC)_{18}.(GC)_{18}$	96	k_{GC}, k_{CG}
$(AC)_{18}.(GT)_{18}$	67	k_{AC}, k_{CA}
$(AG)_{18}.(CT)_{18}$	59	k_{AG}, k_{GA}
$(A)_{36}.(T)_{36}$	45	k_{AA}
$(AGC)_{12}.(GCT)_{12}$	75	k_{AG}, k_{GC}, k_{CA}
poly(AT).poly(AT)[a]	38[a]	k_{AT}, k_{TA}

[a]T_m taken from Ref. (18).

DNA melting curves

All DNA oligos were synthesized and gel- and HPLC-purified at the Keck DNA Synthesis Facility at Yale University. The DNA was dissolved to 200 mM in 30 mM Na phosphate buffer pH 7.5, 100 mM NaCl, 1 mM EDTA. The complementary oligos were annealed by heating to 100°C and stepwise cooling to room temperature in a PCR cycler. The dsDNA oligos were HPLC purifed on a DEAE-Superose column with a LiCl gradient. The absence of hairpin and G-quartet structures in the dsDNA oligos was verified by native 10% PAGE performed at various temperatures. dsDNA melting curves were collected for 20°C–105°C at 250–280 nm on a Varian Cary 50 Bio UV/Vis spectrometer equipped with a Peltier probe. The temperature was varied by 0.5°C/min. The melting curves were obtained for the maximum absorbance wavelength for each oligo and normalized following the procedure of Breslauer (25). Data were collected from five independent experiments.

PBD simulations

The MC protocol used in this work is similar to what has previously been used (9) to interrogate the melting behavior of the PBD model. Steps were performed using a cutoff value of 10 Å and a strand separation threshold of $y_n \geq 1$ Å, above which the DNA was considered melted. At least 1000 simulations with different initial conditions were conducted. MC simulations (9,12) of the L60B36 sequence and Langevin dynamic simulations (27) of the P5 promoter were conducted as previously described, using the determined here new sequence-dependent stacking term.

RESULTS AND DISCUSSION

The PBD model

The potential energy of the PBD model is

$$H = \sum_{n=1}^{N} \left\{ D_n (e^{-a_n y_n} - 1)^2 + \frac{k}{2}(1 + \rho e^{-\beta(y_n + y_{n-1})})(y_n - y_{n-1})^2 \right\} \quad \mathbf{1}$$

where the sum extends over all N base pairs of the DNA molecule and y_n denotes the relative displacement from equilibrium of the n-th base-pair. The first term in Equation (1) is the Morse potential, which represents the hydrogen bonding of a Watson–Crick complementary base pair. The parameters D_n and a_n in this term denote the nature of the n-th base pair, i.e. A–T or G–C.

The second term is an anharmonic potential representing the aromatic stacking interaction between the n- and (n-1)-th consecutive base pairs, augmented by a coupling constant that depends on the relative displacement of these two bases in a nonlinear fashion. This term is essential for simulating the local constraints of nucleotide motion, which result in long-range cooperative effects (5). The values of the relevant model parameters were initially chosen to reproduce the melting transitions of long homogeneous sequences. Subsequently, the values $\beta = 0.35\,\text{Å}^{-1}$, $D_{AT} = 0.05\,\text{eV}$, $D_{GC} = 0.05\,\text{eV}$, $a_{AT} = 4.2\,\text{Å}^{-1}$ and $a_{GC} = 6.9\,\text{Å}^{-1}$, $\rho = 2$ and $k = 0.025$ eV/Å^2 were optimized by fitting UV melting curves of three short heterogeneous DNA sequences (6). The stacking constant k is sequence-independent and represents an average value of the stacking energy of the different dinucleotides. Here we used DNA melting measurements and published data (18) to derive 10 k-parameter values, corresponding to the 10 dsDNA dinucleotide steps.

DNA melting studies

A pioneering experimental study of homogeneous and periodic DNA oligomers (18), which reported large differences in the melting temperatures of poly(dAdT) versus poly(dA).poly(dT) and poly(dGdC) versus poly(dG).poly(dC) formed the basic inspiration for the present work. To obtain precise measurements suited to our needs, we performed UV absorbance melting measurements of six synthetic DNA oligonucleotides designed to contain simple repeating combinations of different dinucleotide steps (Table 1). The absence of hairpin, G-quartet or other secondary structures in the DNA samples was verified by native polyacrylamide electrophoresis (PAGE) on the samples at various concentration and temperatures. A poly(dAdT) oligo consistently formed hairpin structures even when annealed at high concentrations and at different salt concentrations, and was discarded from the study. For the simulations, we used the poly(dAdT) melting data presented by Wells *et al.* (18). As expected, the poly(dG) DNA displayed unusual bands in the presence of K^+ ions, but was of homogenous composition with the same gel mobility as control dsDNA in the sodium phosphate/EDTA buffer used for the melting studies (data not shown). None of the remaining oligos used for the melting simulations (Table 1) displayed any non-dsDNA structures as judged by the PAGE assay (data not shown). The melting temperatures obtained from UV absorbance measurements are shown in Table 1. Our data are consistent with those reported by others (18) in terms of the order of melting temperatures, with quantitative differences due to the different experimental conditions, including buffer composition and the defined length of the oligos.

Melting temperature simulations

To establish a protocol for simulating the experimental DNA melting data (see Materials and Methods section), we used a recently developed MC method (9,12) with the original PBD average stacking constant, and reproduced the melting temperature of a known heterogeneous DNA sequence (6). Previously, we showed that the melting temperature of a given DNA sequence predicted by PBD MC simulations varies almost linearly with the value of a chosen average stacking constant, while the shape of the melting curve does not significantly depend on this constant (12). These results were used to obtain the stacking constants k_{GG} and k_{AA} for the poly(dG).poly(dC) and poly(dA).poly(dT) oligos. Next, we obtained the average stacking constants for the poly(dAdT), poly(dAdC).poly(dGdT), poly(dAdG).poly(dCdT) and poly(dGdC) oligos by fitting to the experimentally obtained melting temperatures. To extract the individual stacking constants from these average values we relied on the following observation: it can be easily shown for widely separated DNA strands that in the harmonic stacking limit the partition function for any given sequence can be expressed exactly in terms of an average stacking constant

$$k_{\text{average}} = \sqrt[N]{\sum_{i=1}^{N} s_i}$$

where k_i are the individual stacking constants of the dinucleotides present in the sequence. This relationship yields one equation connecting each pair of stacking constants participating in the oligos poly(dAdT), poly(dAdC).poly(dGdT), poly(dAdG).poly(dCdT) and poly(dGdC).

A second equation is derived from of the ratios k_{GA}/k_{AG}, k_{CA}/k_{AC}, k_{AT}/k_{TA}, whose values were obtained from the literature. The proper choice of base stacking constants for a given DNA model is hampered by the complexity of the stacking interaction, and the difficulty of separating the individual contributions to dsDNA stability (28). It should be noted that the choice of constants also depends on the model, as evidenced by a number of seemingly conflicting sets of base stacking energies reported in the literature (28). Given the identical hydrogen bonding patterns, reproducing the large difference in the melting temperatures of the polyGG and polyGC oligos, shown in Figure 1, requires a very weak GG/CC stacking term. As thoroughly described previously (28), there is ample evidence that a GG/CC stack dinucleotide is indeed remarkably unstable compared with GC/CG. To be consistent with this observation, we obtained the necessary stacking force constant ratios from a well-established set of intrinsic stacking energies calculated by Sponer *et al.* (29) from force-field calculations with the empirical AMBER force field (30).

The initial values for the individual PBD stacking constants were calculated using the above two independent relations. These values were further adjusted to fit the experimental melting data using the previously established MC protocol (Materials and Methods section). The final 10 stacking constants are shown in Figure 2.

The uniform stacking PBD was conceived largely as a qualitative model aimed at reproducing the basic characteristics of DNA melting transitions. Here, we introduce a more quanitative model that is capable of predicting melting temperatures for the given, physiologically

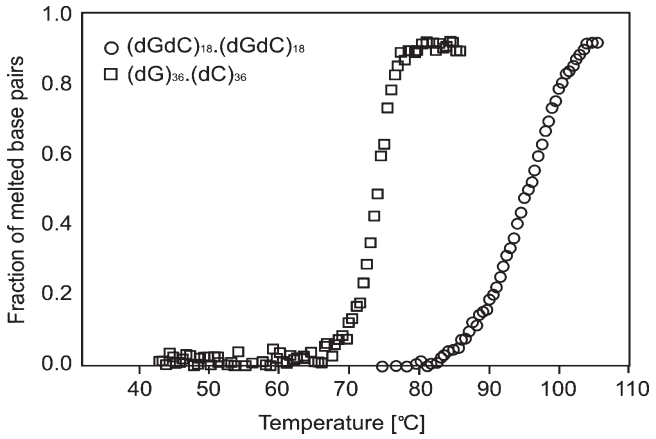

Figure 1. Normalized UV absorption melting curves for the $(dG)_{36}.(dC)_{36}$ (squares) and $(dGdC)_{18}.(dGdC)_{18}$ (circles) oligos.

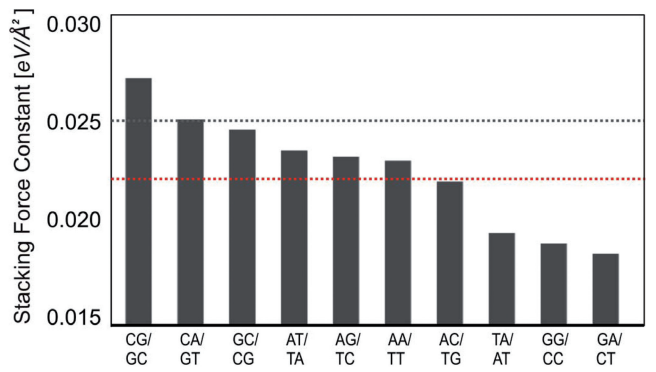

Figure 2. Values of the obtained PBD base stacking force constants in $eV/Å^2$. The identity of the 10 dinucleotide steps is shown below the bars. The stacking force constants are shown as vertical bars. The dashed lines indicate the stacking constant of the average stacking PBD model (black) and the average of the 10 stacking constants of the sequence-dependent PBD model (red).

relevant buffer conditions. The sequence-dependent PBD model has an average stacking constant that is 10% lower (Figure 2) than in the homogenous stacking PBD.

The sequence-dependent stacking force constants were tested on, and accurately reproduced the melting temperature of poly(dAdGdC).poly(dGdCdT) (Figure 3). A comparison of the experimentally determined DNA melting temperatures, and the predictions of various DNA models is shown Figure 3. As can be seen from the figure, the previous PBD model, like most other DNA models performs well for mixed sequences of intermediate G/C content. However, these models fail for sequences of 100% G/C or A/T content. The new sequence-dependent PBD was derived from the shown experimental data (Table 1, Figure 1) and strictly for the given experimental conditions (buffer, salt, DNA concentration). Nevertheless, the advantages of DNA sequence-dependent stacking PBD potential over the previous average stacking potential for thermodynamic and dynamic simulations of genomic sequences are obvious.

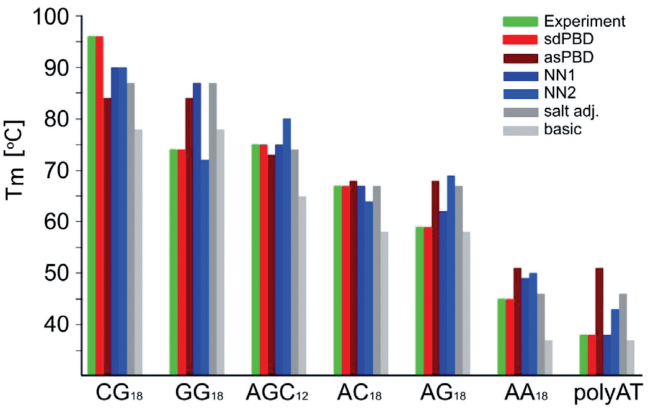

Figure 3. Experimentally determined and calculated melting temperatures of periodic and homogeneous dsDNA sequences. The calculated melting temperatures [T_m (°C), on the vertical axis] are identified as follows: experimentally determined, green bar; sequence-dependent PBD (sdPBD), red; average stacking PBD (asPBD), brown; NN thermodynamic model (NN1, http://www.promega.com/biomath/calc11.htm (16), dark blue; NN thermodynamic model (NN2, http://www.basic.northwestern.edu/biotools/oligocalc.html) (14,31), light blue; salt adjusted model calculated with a regression (http://www.promega.com/biomath/calc11.htm) (32), dark grey; basic regression model (http://www.promega.com/biomath/calc11.htm) (33), light grey. The identity of the sequences is shown bellow the bars.

To further evaluate the performance of our extended PBD model, we repeated simulations previously reported by others. Zeng *et al.* (26) experimentally determined the length and statistical weight of local DNA denaturation bubbles at various temperatures. The PBD model was quite successful at reproducing the nucleation size of internal denaturation bubbles observed in these experiments (9), in contrast to thermodynamic models of DNA melting (26). We repeated the PBD MC simulations (9) with the new sequence-specific stacking constants and found that the two models perform equally well at reproducing the experimental data (data not shown). A more detailed comparison of the two PBD simulations (Figure 4) reveals that the predicted bubble locations are nearly identical. However, the bubble amplitudes predicted by the sequence-dependent model are more pronounced, to various degrees, than the amplitudes predicted by the homogenous stacking model. This effect is independent of the overall baseline shift due to the slightly lower average stacking potential in the new PBD. In this case, the main effect of the intricate cooperativity introduced by the sequence-dependent stacking is larger amplitudes in the bubble-forming region of the DNA sequence.

In addition to the bubble probabilities and amplitudes that can be calculated using PBD MC and thermodynamic methods, Langevin dynamics simulations yield the lifetimes of transient DNA bubbles (B. S. Alexandrov *et al.*, submitted for publication; 27). Here, we repeated previous simulations of the DNA bubble dynamics of the P5 adenoassociated virus core promoter (100 bp) and a transcriptionally silent P5 mutant variant (27). *In vitro* transcription, single strand nuclease data and PBD thermodynamic calculations for this promoter (7,21) have

shown strong correlation between transcriptional activity and the presence of a transient bubble at the transcriptional start site. PBD Langevin simulations suggested that bubbles occurring at the start site are more stable than equally likely but shorter lived bubbles forming at the TATA-box of this promoter (B. S. Alexandrov *et al.*, submitted for publication; 27). A propensity of the P5 start site to open transiently was also suggested by *in vitro* transcription experiments, in which negative supercoiling of a circular DNA template was equivalent to introduction of a 5 bp mismatch at the start site of a linearized template (21). Figure 5 shows bubble lifetimes calculated from PBD Langevin dynamic trajectories of the P5 promoter with the sequence-dependent stacking potential. The simulations predict a stable bubble at the P5 start site, and a lack of dynamic activity at the same location in the transcriptionally silent mutant. The recalculated, with the new stacking, bubble life times, for the P5—wild-type and P5 mutant promoters, showed more delineated differences than previously reported (27).

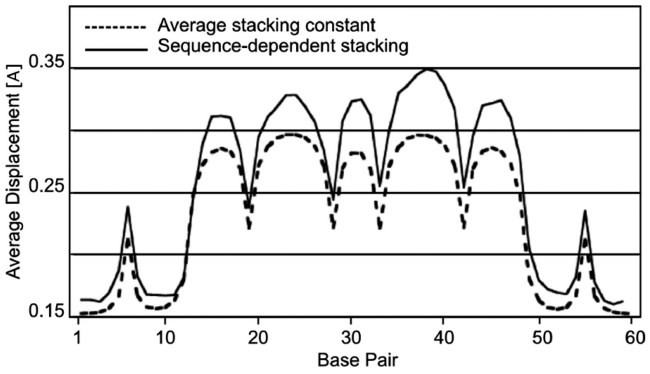

Figure 4. Average strand displacements for the sequence L60B36, calculated by MC PBD simulations with an average (dashed line) and a sequence-dependent (solid line) stacking term. Strand displacements are presented on the vertical axis in Å. The base pare position is shown on the horizontal where base pair 1 is the first base at the 5′ end of the sequence.

CONCLUSION

We have obtained 10 base stacking force constants that systematically improve the performance of the PBD model for homogenous and periodic DNA sequences. The new stacking constants are all within 20% of the previously used average stacking constant, suggesting that previous PBD calculations on mixed DNA sequences are reasonable. The experiments and numerical simulations reported here show that the GG/CC dinucleotide stacking is remarkably unstable, compared with the GC/CG and CG/GC dinucleotide steps. Similar experimental results (18) and *ab initio* and force field calculations (28,29) have been reported, but have been frequently overlooked due to the strength of the G–C hydrogen bond, which dominates as a determinant of dsDNA stability. However, such differences in the base stacking interactions must contribute to the dynamic behavior of G/C-rich genomic sites, with likely relevance to genomic function.

The modified PBD model demonstrates significantly improved performance in predicting melting temperatures for all tested homogeneous and periodic DNA sequences, regardless of G/C content.

As we previously reported, PBD dynamic and thermodynamic calculations predict the existence of transient thermally induced bubbles at the transcriptional start sites of many eukaryotic promoters. Langevin dynamic simulations with the new PBD model on the A/T-rich viral P5 promoter strongly support these predictions.

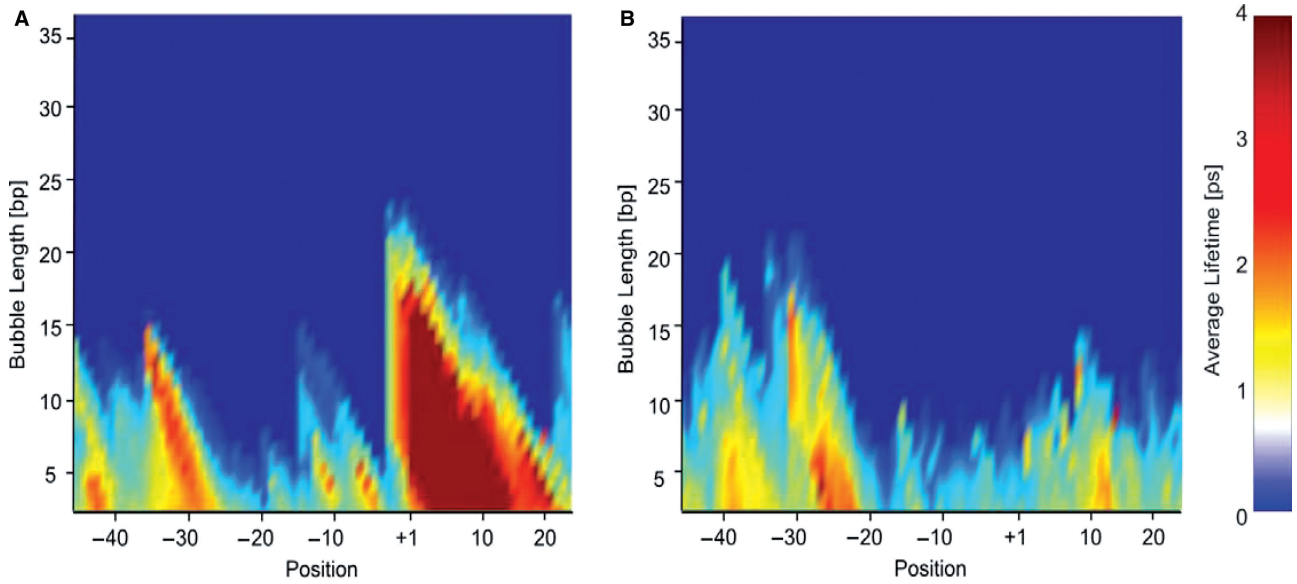

Figure 5. Average bubble times (color scales) for (**A**) P5-wild type and (**B**) P5-mutant promoters. The bubble times are given in picoseconds on the color scale. The bubble length is shown on the vertical axis in units of base pairs. The positions of the base pairs are shown relative to the transcription start site (+1) on the horizontal axis.

FUNDING

National Institutes of Health (GM073911 to A.U.); National Nuclear Security Administration of the US Department of Energy at Los Alamos National Laboratory (Contract DE-AC52-06NA25396). Funding for open access charge: DE-AC52-06NA25396 and GM073911.

Conflict of interest statement. None declared.

REFERENCES

1. McClare,C.W. (1972) A quantum mechanical muscle model. *Nature*, **240**, 88–90.
2. Davydov,A.S. (1973) The theory of contraction of proteins under their excitation. *J. Theor. Biol.*, **38**, 559–569.
3. Englander,S.W., Kallenbach,N.R., Heeger,A.J., Krumhansl,J.A. and Litwin,A. (1980) Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. *Proc. Natl Acad. Sci. USA*, **77**, 7222–7226.
4. Peyrard,M. and Bishop,A.R. (1989) Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, **62**, 2755–2758.
5. Dauxois,T., Peyrard,M. and Bishop,A.R. (1993) Entropy-driven DNA denaturation. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **47**, R44–R47.
6. Campa,A. and Giansanti,A. (1998) Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains. *Phys. Rev. E.*, **58**, 3585.
7. Choi,C.H., Kalosakas,G., Rasmussen,K.O., Hiromura,M., Bishop,A.R. and Usheva,A. (2004) DNA dynamically directs its own transcription initiation. *Nucleic Acids Res.*, **32**, 1584–1590.
8. Kalosakas,G., Rasmussen,K., Bishop,A.R., Choi,C.H. and Usheva,A. (2004) Sequence-specific thermal fluctuations identify start sites for DNA transcription. *Eur. Phys. Lett.*, **68**, 127.
9. Ares,S., Voulgarakis,N.K., Rasmussen,K.Ø. and Bishop,A.R. (2005) Bubble nucleation and cooperativity in DNA melting. *Phys. Rev. Lett.*, **94**, 035504.
10. Voulgarakis,N.K., Redondo,A., Bishop,A.R. and Rasmussen,K.Ø. (2006) Probing the mechanical unzipping of DNA. *Phys. Rev. Lett.*, **96**, 248101.
11. Voulgarakis,N.K., Kalosakas,G., Rasmussen,K. and Bishop,A.R. (2004) Temperature-dependent signatures of coherent vibrational openings in DNA. *Nano Lett.*, **4**, 629.
12. Alexandrov,B.S., Voulgarakis,N., Rasmussen,K.Ø., Usheva,A. and Bishop,A.R. (2009) Pre-melting dynamics of DNA and its relation to specific functions. In Special issues: DNA melting. *J. Phys. Condens. Matter*, (in press).
13. Choi,C.H., Rapti,Z., Gelev,V., Hacker,M.R., Alexandrov,B., Park,E.J., Park,J.S., Horikoshi,N., Smerzi,A., Rasmussen,K.Ø. *et al.* (2008) Profiling the thermodynamic softness of adenoviral promoters. *Biophys. J.*, **95**, 597–608.
14. Breslauer,K.J., Frank,V., Blöcker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.
15. SantaLucia,J., Allawi,H.T. and Seneviratne,P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.
16. SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
17. Blake,R.D., Bizzaro,J.W., Blake,J.D., Day,G.R., Delcourt,S.G., Knowles,J., Marx,K.A. and SantaLucia,J. Jr (1999) Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics*, **15**, 370–375.
18. Wells,R.D., Larson,J.E., Grant,R.C., Shortle,B.E. and Cantor,C.R. (1970) Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences. *J. Mol. Biol.*, **54**, 465–497.
19. Poland,D. and Sheraga,H.A. (1966) Phase transitions in one-dimension and the helix-coil transition in polyamino acids. *J. Chem. Physics*, **45**, 1456–1463.
20. Yakovchuk,P., Protozanova,E. and Frank-Kamenetskii,M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.*, **34**, 564–574.
21. Usheva,A. and Shenk,T. (1996) YY1 transcriptional initiator: protein interactions and association with a DNA site containing unpaired strands. *Proc. Natl Acad. Sci. USA*, **93**, 13571–13576.
22. Leblanc,B.P., Benham,C.J. and Clark,D.J. (2000) An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled. *Proc. Natl Acad. Sci. USA*, **97**, 10745–10750.
23. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
24. Goldstein,R.F. and Benight,A.S. (1992) How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers*, **32**, 1679–1693.
25. Breslauer,K.J. (1995) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Meth. Enzymol.*, **259**, 221–242.
26. Zeng,Y., Montrichok,A. and Zocchi,G. (2004) Bubble nucleation and cooperativity in DNA melting. *J. Mol. Biol.*, **339**, 67–75.
27. Alexandrov,B.S., Wille,L.T., Rasmussen,K.Ø., Bishop,A.R. and Blagoev,K.B. (2006) Bubble statistics and dynamics in double-stranded DNA. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **74**, 050901.
28. Sponer,J., Riley,K.E. and Hobza,P. (2008) Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.*, **10**, 2595–2610.
29. Sponer,J., Jurecka,P., Marchan,I., Luque,F.J., Orozco,M. and Hobza,P. (2006) Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem. Eur. J.*, **12**, 2854–2865.
30. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M. Jr, Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
31. Kibbe,W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*, **35**, W43–W46.
32. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33. Rychlik,W. and Rhoads,R.E. (1989) A computer program for choosing optimal oligonudeotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.*, **17**, 8543.