Published online March 5, 2004

1584–1590 *Nucleic Acids Research, 2004, Vol. 32, No. 4*
DOI: 10.1093/nar/gkh335

DNA dynamically directs its own transcription initiation

Chu H. Choi, George Kalosakas[1], Kim Ø. Rasmussen[1], Makoto Hiromura, Alan R. Bishop[1] and Anny Usheva*

Endocrinology, Beth Israel Deaconess Medical Center and Harvard Medical School, Department of Medicine, 99 Brookline Avenue, Boston, MA 02215, USA and [1]Center for Nonlinear Studies and Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received January 16, 2004; Revised and Accepted February 19, 2004

ABSTRACT

It has long been known that double-stranded DNA is subject to temporary, localized openings of its two strands. Particular regions along a DNA polymer are destabilized structurally by available thermal energy in the system. The localized sequence of DNA determines the physical properties of a stretch of DNA, and that in turn determines the opening profile of that DNA fragment. We show that the Peyrard–Bishop nonlinear dynamical model of DNA, which has been used to simulate denaturation of short DNA fragments, gives an accurate representation of the instability profile of a defined sequence of DNA, as verified using S1 nuclease cleavage assays. By comparing results for a non-promoter DNA fragment, the adenovirus major late promoter, the adeno-associated viral P5 promoter and a known P5 mutant promoter that is inactive for transcription, we show that the predicted openings correlate almost exactly with the promoter transcriptional start sites and major regulatory sites. Physicists have speculated that localized melting of DNA might play a role in gene transcription and other processes. Our data link sequence-dependent opening behavior in DNA to transcriptional activity for the first time.

INTRODUCTION

Genetic information is stored in the sequence of DNA bases, which are protectively embedded in the hydrophobic center of the distinctive double helix. In order to read this code, the bases must be physically exposed by locally separating the strands at the transcriptional start site (TSS). This feature of transcription initiation is rigorously conserved, from prokaryotes to humans. The denaturation of double-stranded DNA (dsDNA) in the correct places is essential for initiation of specific gene transcription.

There is good evidence that torsional effects play a significant role in opening the two strands, through stress-induced duplex destabilization (SIDD) (1). Supercoiling of dsDNA clearly alters the profile of the free energy associated with the local melting of the double helix at each base pair. Using computational models, it has been predicted that specific sequences will open in places known to be important in initiating and regulating transcription (2,3). While SIDD may well represent the mechanism of localized DNA melting, it is not entirely clear how the specificity of the position of opening is realized in dsDNA. In order to understand the fundamental selectivity of DNA opening, it is important to isolate the physical properties associated with a particular DNA sequence from these synergistic torsional effects.

Because adenine–thymine and guanine–cytosine base pairs differ in their hydrogen bonding and stacking, the local DNA sequence is critical to the physical properties of a stretch of DNA, affecting its melting temperature (4), helical structure (5) and elasticity (6), among other parameters. But in the search for methods of promoter prediction, it has become obvious that consensus sequence alone is not sufficient to determine transcriptional activity of a particular DNA sequence (7). If torsional strain is primarily responsible for strand separation, openings would be expected to form exclusively on AT-rich stretches of DNA, which contain fewer hydrogen bonds than GC base pairs. While opening sites on DNA are often AT-rich, it is plainly not the case that they always represent the best opening sites in dsDNA (2,8). Rather, the physical and molecular properties of a given sequence of DNA define its complex behavior.

Thermally induced breathing of the two DNA strands is a well-documented phenomenon, and can result in structures significantly different from equilibrium helices (9,10). It can give rise to localized DNA melting of 10 bp or more at physiological temperatures, which is similar in size to some transcriptional bubbles (11–13). Theoretical studies have predicted sequence-dependent localized hot spots, or partial denaturations of the double helix, to form with the available thermal energy at temperatures as low as 50–70K (14,15). These DNA openings occur spontaneously with thermal energy, demonstrating a crucial role for entropy. The pattern

*To whom correspondence should be addressed. Tel: +1 617 632 0522; Fax: +1 617 632 2927; Email: ausheva@bidmc.harvard.edu
Present address:
Makoto Hiromura, Division of Cancer Biology and Institute for Genetic Medicine, Hokkaido University, N15, W7, Kita-ku, Sapporo 060-0815, Japan

of these relatively flexible regions of DNA is dependent upon localized sugar–phosphate backbone torsion, base pair hydrogen bonding and adjacent π-electron interactions. Thus, the DNA sequence imposes a colored spatial and temporal stochastic environment, with preferential sites of opening.

The inherent tendency of a given base pair to open is determined by its sequence-specific context. Opening of any base pair is coupled to its neighbors, resulting in a cooperative effect which is observed experimentally (16). Therefore, localized openings in DNA can nucleate larger openings. It is reasonable to assume that some fluctuational openings in dsDNA may, through the aid of supercoiling or other means, grow to become comparable in size to the transcription bubble at physiologically relevant temperatures. Indeed, supercoiling has been associated with increased breathing behavior over the relaxed form of DNA (17). The formation of these temporary single-stranded regions in relaxed DNA might nucleate significant molecular events, such as the initiation of transcription (18,19). To test this hypothesis, we examined a non-promoter DNA sequence from the cDNA of human transcription factor IIB and two well-characterized transcriptionally active promoters: the adenovirus major late promoter (AdMLP) and the adeno-associated viral P5 promoter (AAV P5).

MATERIALS AND METHODS

Non-promoter DNA sequence

We used a synthetic 62 bp fragment of the cDNA duplex from the human general transcription factor IIB as a non-promoter control. This fragment is within the coding region of the gene and has no known activity for transcription initiation. These nucleotides correspond with positions 738–799 of NM001514.

Nonlinear DNA dynamical model

Langevin molecular dynamics simulations were applied to dsDNA sequences using the Peyrard–Bishop (PB) model (20,21). The sequence was repeated on both ends of the fragment to avoid terminal base pair effects, effectively circularizing the DNA sequence without any torsional effects. Simulations were run on several Sun Blade workstations. 100 separate PB model realizations were simulated over a 1 ns timescale using 1 fs intervals, yielding 10^8 data points per sequence. Instances of a 10 bp opening over a threshold value of 2.1 Å were recorded, and the values at the central base pair are reported.

In vitro transcription assays

In vitro run-off transcription assays were used to demonstrate specific polymerase II transcription. Reactions were assembled at 28°C with HeLa nuclear extract as previously described (22). A 105 bp linear DNA fragment with a partial sequence corresponding to the human transcription factor IIB was used as a non-promoter control template. A linear DNA fragment of 120 nt containing the AdMLP or P5 promoter was used as the DNA template. The reaction products were monitored by run-off transcription in the presence of [^{32}P]UTP. The reaction products were separated by gel electrophoresis and visualized by autoradiography of the ^{32}P-labeled transcripts. A Molecular Dynamics Phosphorimager 400-B was used to document the results.

S1 nuclease cleavage assays

The non-coding (lower) strand for the non-promoter control and the P5 promoter, and the coding (upper) strand for the AdMLP was ^{32}P-labeled at the 5′-terminus with T4 polynucleotide kinase (Invitrogen). The labeled and the unlabeled strands were then annealed by temperature cycling, and the dsDNA was used as a substrate for S1 nuclease cleavage. 0.2 nM dsDNA was incubated with S1 nuclease (50 U of enzyme per reaction) at 28°C for 45 min in buffer containing 10 mM HEPES (pH 7.2), 50 mM NaCl and 4 mM $Zn(C_2H_3O_2)_2$, as recommended by the supplier (Roche). The reaction was stopped with 20 mM EDTA. After ethanol precipitation, the DNA digestion products were electrophoretically separated on a 10% sequencing gel (National Diagnostics). A Molecular Dynamics PhosphorImager 400-B was used to document the results.

RESULTS

Our data give credence to the formation of temporal, localized thermal openings in double-stranded promoter DNA, but not in a non-promoter control DNA duplex. Both our theoretical simulations and our cleavage assays using S1 nuclease support preferential openings are consistent with transcriptionally important sites in promoter DNA, suggesting that the thermal opening profile of DNA may play a role in transcription initiation.

Non-promoter DNA sequence

We used a 62 bp DNA fragment amplified from the cDNA of human transcription factor IIB (23) as a non-promoter control sequence (Fig. 1A). This DNA sequence is free from any known protein interaction sites. By examining a 105 bp fragment containing this sequence, we verified that it contains no transcriptional activity *in vitro*. Runoff transcription reactions on this linear template resulted in only free [^{32}P]UTP (Fig. 1B, lane 2).

To simulate localized melting within the DNA fragments, we applied the PB model (20,21). In the DNA adaptation of the PB model, semi-empirical Morse potentials represent the hydrogen bonds of a GC or AT base pair, and consecutive base pairs are coupled through a nonlinear term that represents vertical π-stacking interactions. These explicit definitions of interactions for each base distinguish our simulations from those of Benham (24), yielding a much more sensitive model, at greater computational cost. While SIDD calculations can determine denaturation stretches in the order of 100 bp in kilobase DNA sequences, our model can predict the fluctuational propensity of one specific base pair out of a 100 base sequence. The central feature of the PB model is the entropy-driven nucleation of coherent conformational openings in dsDNA, which ultimately result in DNA melting (16,21). PB simulations have shown remarkable agreement with experimental melting curves of short dsDNA chains (25), though the method does not explicitly account for salt ions or water molecules. These effects are absorbed into the semi-empirical parameters and do not require explicit definition in a correct modeling of the nonlinearities of the system.

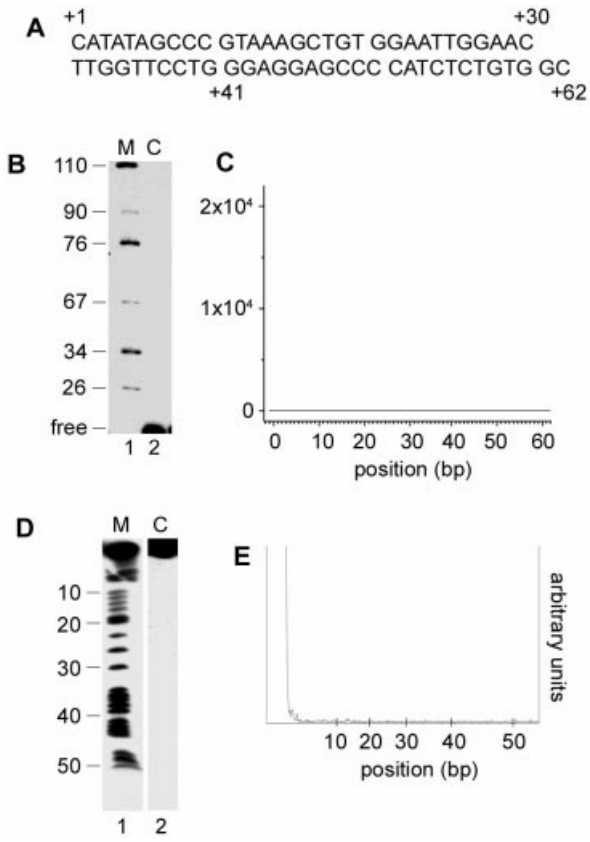

Figure 1. Analysis of non-promoter control DNA fragment. (**A**) Upper strand sequence of the 62 bp non-promoter DNA. (**B**) Run-off transcription assay on a 105 bp linear fragment containing the non-promoter sequence in nuclear extract. The migration position of marker in base pairs and the position of free [^{32}P]UTP is shown to the left of the panel. Lane 1, DNA marker (M); lane 2, reaction products (C). (**C**) PB simulation of the non-promoter sequence, plotting simulated instances of 2.1 Å-separated openings of 10 bp or more versus base position in the sequence. (**D**) S1 nuclease cleavage assay of the non-promoter DNA fragment. The corresponding sequence position is indicated to the left of the panel. Lane 1, lower strand-labeled P5 promoter GA sequencing reaction was used as a marker (M); lane 2, non-promoter DNA S1 digestion reaction (C). (**E**) Cleavage density profile of the non-promoter DNA sequence in the S1 nuclease experiment.

Computational results for our control sequence show almost no tendency to form 10 bp openings by the 62 bp DNA duplex (Fig. 1C). This result does not negate the possibility of the formation of smaller destabilizations, or 10 bp openings of smaller magnitude. This sequence has a GC content of 54.84%, with these triply hydrogen bonded base pairs fairly evenly spaced apart. As a result, the plot of the opening profile shows negligible opening along the whole DNA sequence.

S1 nuclease is a member of a family of glycoprotein nucleases which has been shown to selectively cleave single-stranded DNA (26). It has been widely employed in studies of transcriptional systems. While it is sufficiently sensitive to react with arrested transcriptional elongation complexes, which feature a relatively large transcriptional 'bubble', the physical size of S1 nuclease reduces its ability to detect smaller temporal openings in dsDNA. The bulkiness of S1 nuclease aids in selectively cleaving larger temporary openings over small openings, but it results in a very weak signal. We overcame this issue of sensitivity by incubating the dsDNA with S1 nuclease for a longer period of time than in usual assays (45 min). S1 nuclease displays optimal activity under low pH conditions, but it is noteworthy that it retains its reactivity even at pH 7.2 in our experiments.

Cleavage of the upper-strand labeled control sequence dsDNA fragment yields no detectable S1 nuclease activity, even after extended periods of incubation (Fig. 1D, lane 2). The top of the lane reflects the large amount of uncleaved template in the sample, and should not be taken as an opening peak. The density profile for the control lane shows not even slight cleavage under these conditions (Fig. 1E). This agrees with the prediction that we do not expect the formation of large (10 bp or greater) denaturations along this particular DNA sequence.

The AdMLP

The AdMLP regulates the adenoviral late gene transcription (27). AdMLP has proven useful as a model system for studying basic and regulated TATA box-binding protein (TBP)-dependent gene transcription (28). Extensive theoretical and experimental work has shown the importance of dynamics and flexibility of the TATA box in DNA (29–32). We examined the 86 bp major late core promoter (Fig. 2A), which includes the known TSS, labeled '+1'. This promoter sequence is sufficient to direct specific transcription initiation in nuclear extracts (Fig. 2B).

The computational simulation results for the AdMLP sequence reveal three peaks marking the preferential positions of opening in this DNA fragment (Fig. 2C). In our simulations, we have constrained the opening size to be at least 10 bp in length. These constraints limit our results to DNA openings appropriately sized for transcription. Despite the fact that the AdMLP sequence contains a relatively high GC content of 66.28% compared to the control sequence, it yielded peaks with an intensity of nearly 10^4 instances out of 10^8 data points. The largest predicted position of opening appears in the region just downstream of the +1 site, where transcription occurs. It is very interesting that the dominant position for opening along the promoter occurs at the same place where transcription is to begin. Another slightly smaller peak is centered about the –29 bp position, and a minor peak appears at position –9. The –29 location on the AdMLP corresponds to the TATA box, the known binding site for TBP (33). The TATA box is an AT-rich sequence of DNA which is required for TBP-dependent transcription from this promoter.

The AdMLP is predicted by our model to open preferentially in these two significant places of protein interaction. This is in marked contrast to the PB simulation results for the non-promoter sequence, which has no known function as a gene promoter or regulation site. The sharp differences in the opening profiles of AdMLP and the negative control DNA suggest that high levels of thermal opening may be characteristic of gene promoters but not other DNA sequences.

Digestion of the radioactively labeled upper strand of the AdMLP duplex with S1 nuclease results in cleavage of four main regions of the oligonucleotide, centering around positions +6, +1, –10 and –31 (Fig. 2D, lane 2). The locations of these opening sites are in excellent agreement with the simulation data (Fig. 2E). The observed template cleavage at +1 strongly correlates with the TSS, and the opening centered at –31 should represent the DNA elasticity at the TATA box.

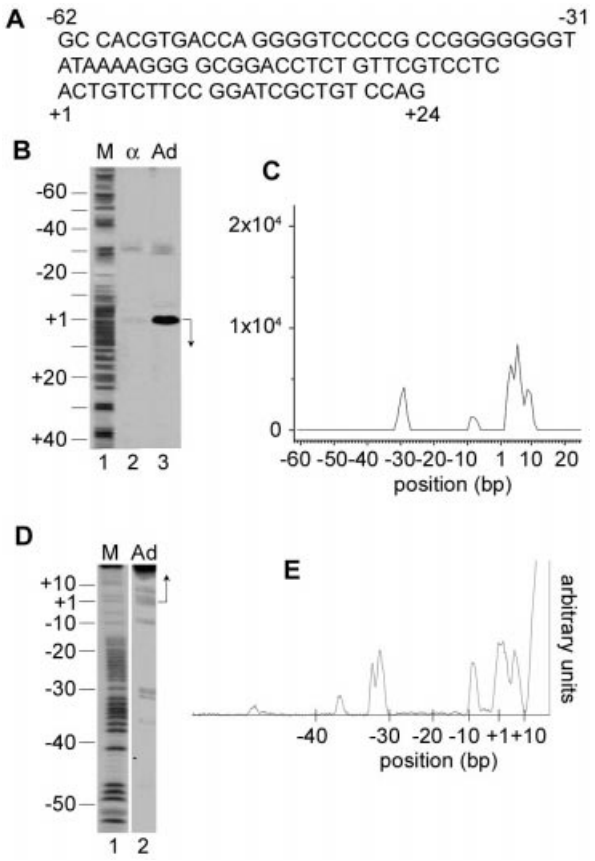

Figure 2. Analysis of adenovirus major late promoter. (**A**) Upper strand sequence of the 86 bp AdMLP. (**B**) Run-off transcription assay in nuclear extract using a 120 bp linear template containing the 86 bp fragment of the AdMLP. The arrow on the right indicates the TSS and the direction of transcription. The corresponding sequence position is indicated to the left of the marker. Lane 1, GA DNA sequencing reaction was used as a marker (M); lane 2, transcription with α-amanitin to specifically arrest polymerase II transcription (α); lane 3, RNA transcription product (Ad). (**C**) PB simulation of the AdMLP sequence, plotting simulated instances of 2.1 Å-separated openings of 10 bp or more versus base position in the sequence. (**D**) S1 nuclease cleavage assay of the AdMLP. The corresponding sequence position is indicated to the left of the marker. Lane 1, upper strand-labeled AdMLP fragment sequencing reaction was used as a marker (M); lane 2, S1 digestion of the ^{32}P-labeled AdMLP fragment (Ad). (**E**) Cleavage density profile of the AdMLP fragment in the S1 nuclease experiment.

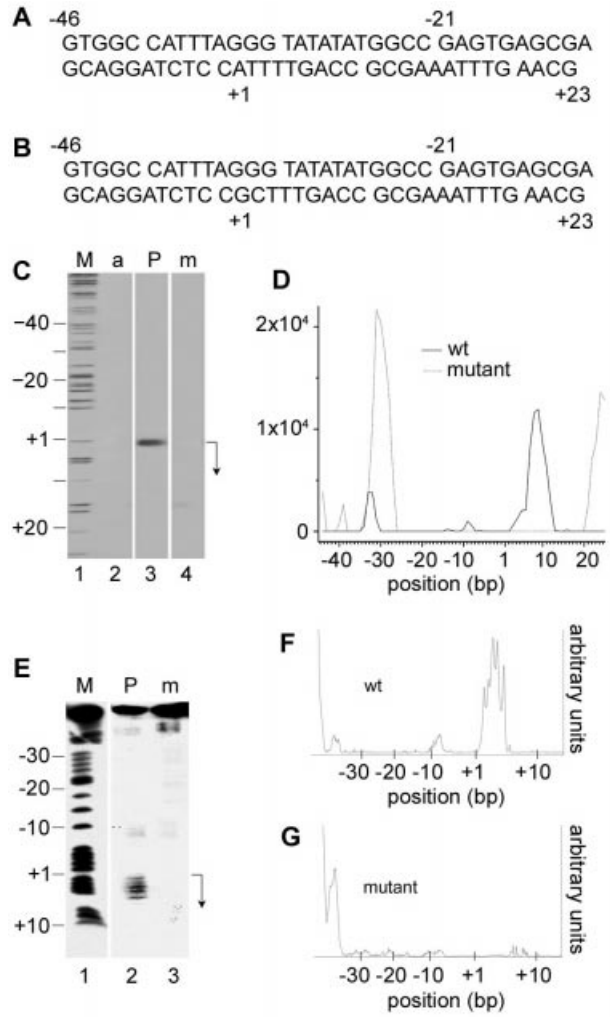

Figure 3. Analysis of AAV P5 promoter and P5 mutant promoter. (**A**) Upper strand sequence of the 69 bp P5 core promoter. (**B**) Upper strand sequence of the 69 bp mutant P5 promoter. (**C**) Transcription assay on a 120 bp fragment containing the P5 promoter and P5 mutant fragment in nuclear extract. The arrow on the right indicates the TSS and the direction of transcription. The corresponding sequence position is indicated to the left of the marker. Lane 1, GA DNA sequencing reaction was used as a marker (M); lane 2, transcription from the P5 promoter with α-amanitin (a); lane 3, RNA transcription products with P5 promoter (P); lane 4, RNA transcription products from the P5 mutant template (m). (**D**) PB simulation of the AAV P5 sequence, plotting simulated instances of 2.1 Å-separated openings of 10 bp or more versus base position in the sequence. The solid line represents the results of the wild-type (wt) P5 promoter, and the broken line represents results with the mutant P5 promoter sequence. (**E**) S1 nuclease cleavage of the P5 promoter and the P5 mutant promoter. The corresponding sequence position is indicated to the left of the panel. Lane 1, lower strand-labeled P5 promoter GA sequencing reaction (M); lane 2, AAV P5 promoter S1 cleavage reaction (P); lane 3, AAV P5 mutant promoter cleavage reaction (m). (**F**) Cleavage density profile of the wild-type (wt) AAV P5 promoter DNA in the S1 nuclease experiment. (**G**) Cleavage density profile of the mutant AAV P5 promoter DNA in the S1 nuclease experiment.

The weak band at –9 was also predicted by the PB simulations. It has been shown that bending of DNA at the TATA box upon binding of TBP is dependent on the local DNA sequence (34). This suggests that altering the sequence does impact the local physical properties of DNA.

The AAV P5 promoter

The AAV P5 promoter (Fig. 3A) directs the expression of Rep78 and Rep68, which are essential for AAV DNA replication (35) and the regulation of AAV gene expression (36). The P5 promoter has been shown to bind the transcription factor Yin Yang 1 (YY1) and to be active for TBP-independent transcription (22,37,38). We verified this specific transcription using reconstituted transcription reactions (Fig. 3C, lane 3). In this study, we examine the 69 bp DNA core promoter, which includes the known TSS.

The P5 promoter has a GC content of only 49.28%, much lower than the AdMLP, but simulations suggest that the two promoters are rather similar. In the 69 bp P5 promoter sequence, computer simulations suggest three preferential sites of opening, centered at positions +9, –9 and –33 (Fig. 3D, solid line). The dominant peak is again located somewhat near

the TSS, although it is significantly more downstream than in the case of AdMLP. Perhaps this difference forms a basis for the difference between TBP-dependent and TBP-independent transcription initiation. The intensity of the largest peak, in the order of 10^4 instances, is very similar to that found for the AdMLP, reinforcing the link between promoter activity and temporal DNA openings. The predicted peak at the –33 position corresponds roughly to a second strong binding site for YY1 (39). This site has been implicated in the regulation of transcription on this promoter. Again, the intensities of peaks in this promoter contrast sharply with those of the non-promoter control sequence, which showed no significant openings in our simulations.

Digestion of a radioactively labeled lower strand of the P5 core promoter duplex with S1 nuclease yields cleavage in three localized regions of the DNA sequence, at the +5, –8 and –36 positions (Fig. 3E, lane 2). Again we note the close correlation of these experimental results with the predicted simulation peaks (Fig. 3F). The band observed at +5 is close to the major opening predicted by our method, and shares the property of being well downstream of the TSS. The opening at –36 correlates with the second YY1 binding site, while the minor opening at –8 matches up well with the simulation peaks. Not only do these experimental results align well with the simulation data, but there is a clear similarity between the experimental breathing profiles of the P5 promoter and the AdMLP, both in peak locations and intensities.

In order to better characterize the effects of sequence on the localized melting of DNA, we also examined a mutant variant of the P5 promoter (Fig. 3B). In this promoter sequence, the two bases at positions +1 and +2 (AT) have been changed to GC on the coding strand, along with the appropriate complements on the lower strand. This mutant promoter no longer binds the YY1 protein at the +1 site, and is completely inactive for transcription (Fig. 3C, lane 4).

Examination of the computed opening profile for this promoter shows that the peak near the TSS in the wild-type P5 promoter has disappeared in conjunction with the transcriptional activity, leaving the second peak near the –32 position intact (Fig. 3D, broken line). The large changes in the vicinity of the TSS brought about by the replacement of just two adjacent base pairs suggests that the opening profile is largely dependent on short range effects, on the order of just a few base pairs. The dominant peak at position –32 has actually increased in magnitude over the corresponding peak in the wild-type promoter. This reflects the result of removing the dominant opening near the TSS. By effectively changing the opening potential energy at one end of the duplex fragment, competition for the available thermal energy is lessened and the opening at position –32 increases its probability of formation. This result underscores the complex coupling of DNA, even over relatively large distances. Though single openings can be efficiently modulated by changing just a few local base pairs, other unintended effects may result elsewhere along the DNA sequence.

The S1 nuclease assay also affirms that the propensity of the mutant promoter to open just downstream of the TSS is undetectable under the reaction conditions (Fig. 3E, lane 3). The lack of a viable site of opening near the TSS corresponds with the loss of transcriptional activity for this promoter, again reinforcing the link between localized opening profiles and significant sites in DNA sequences (Fig. 3G). The observed peak at position –36 has grown significantly in intensity, in accord with PB model predictions.

DISCUSSION

From the theoretical and experimental results given here, it is clear that promoter DNA exhibits a sequence-dependent (but not sequence-specific) predilection to open at or near the TSS. This is demonstrated for both a TBP-dependent and a TBP-independent transcriptional system. The P5 promoter, however, exhibits its major opening farther downstream than does the AdMLP. This difference may provide a basis for discriminating between TBP-dependent and TBP-independent transcription. Both promoter sequences tested showed similar opening profiles, a pattern which may be found in other promoter systems. In contrast, a 62 bp stretch of DNA without transcriptional activity showed no preferred sites for large (10 bp) openings, and a mutant P5 sequence inactive for transcription also lacked a site of opening near the TSS.

The tendency for promoter DNA for localized temporal melting, or pre-melting, is independent of any DNA manipulation, protein interactions or regulation, and so is contained within the DNA sequence itself. This finding indicates that the sequence dependence of DNA opening is not simply a matter of sequence-identity, but an interplay of interactions in determining the physical properties of a stretch of DNA. Transcriptional elements such as the TATA box may not be sufficient to predict localized, transcriptionally-relevant opening formation.

It is also interesting to note that a significant tendency to open is found at known binding sites for transcription factors and regulatory proteins on our promoters. Perhaps this suggests a mode of protein–DNA recognition which is based not directly on sequence, but the formation of localized DNA openings. This may rationalize partial non-specific binding of DNA by proteins (40) and the stochastic binding and dissociation of transcription factors from a particular DNA template.

Breathing normal modes are present in all molecules, and are a result of non-zero thermal energy. This energy is manifested through vibrational modes, on a small spatial scale within covalent bonds, and on a larger scale in weaker hydrogen bonds. It is important to distinguish the temporal DNA openings in our investigation from transcriptional bubble formation or DNA remodeling dynamics (10). Localized melting is on a very slight scale, with a shift of only 1 or 2 Å and, though the probability of opening formation is small, coupled molecular motion occurs on a picosecond to nanosecond timescale. This is distinct from other DNA dynamical features, which occur on a millisecond to microsecond timescale, involve much larger rearrangements and require large inputs of energy in order to appear.

Both theoretical (41) and experimental studies (8) have indicated that, once formed, temporal openings in dsDNA have lifetimes of the order of 1 μs, a comparatively long period of time in relation to molecular diffusion rates. This should allow enough time for a slightly open DNA helix possibly to undergo molecular binding events or other modifications. Though the thermally-induced strand separation is only of the order of 2 Å, the stretched hydrogen bonds

in a given base pair are severely weakened, and nonlinear cooperative effects can serve to lower the barrier even further to opening of the strands. In accordance with this view, it appears that the rate of transcription is governed by the kinetics of protein–promoter DNA binding (42). These results underscore the importance of dynamics in the interpretation of biological and chemical events. The static, average structure of a molecule is not sufficient for function, but it is the correct temporary conformation and orientation which allow a reaction to occur. This is an explicit example of the conformational energy landscapes concept (43).

Using dynamics simulations, we were able to consistently predict preferential opening sites on DNA sequences, and show their correlation to transcriptional start sites and regulatory sites. Our experimental data verify our predicted results in the cases examined. Whether DNA dynamics can be used to determine transcriptionally significant sites on a much larger scale remains to be seen. This study illustrates that our conception of DNA should evolve to include dynamic structural fluctuations due to latent thermal energy. Coherent localized dynamics of DNA may yet prove to be a major determinant of DNA function.

ACKNOWLEDGEMENTS

This work was supported by NIH grant HL624458 to A.U. and by the US Department of Energy under contract W-7405-ENG-36.

REFERENCES

1. Benham,C.J. (1992) Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.*, **225**, 835–847.
2. Benham,C.J. (1993) Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl Acad. Sci. USA*, **90**, 2999–3003.
3. Benham,C.J. (1996) Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J. Mol. Biol.*, **255**, 425–434.
4. Breslauer,K.J., Frank,R., Blocker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.
5. Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, **295**, 85–103.
6. Lankas,F., Sponer,J., Hobza,P. and Langowski,J. (2000) Sequence-dependent elastic properties of DNA. *J. Mol. Biol.*, **299**, 695–709.
7. Pedersen,A.G., Baldi,P., Chauvin,Y. and Brunak,S. (1999) The biology of eukaryotic promoter prediction—a review. *Comput. Chem.*, **23**, 191–207.
8. Dornberger,U., Leijon,M. and Fritzsche,H. (1999) High base pair opening rates in tracts of GC base pairs. *J. Biol. Chem.*, **274**, 6957–6962.
9. Kearns,D.R. (1984) NMR studies of conformational states and dynamics of DNA. *CRC Crit. Rev. Biochem.*, **15**, 237–290.
10. Levitt,M. (1983) Computer simulation of DNA double-helix dynamics. *Cold Spring Harb. Symp. Quant. Biol.*, **47** (Pt 1), 251–262.
11. Korzheva,N., Mustaev,A., Kozlov,M., Malhotra,A., Nikiforov,V., Goldfarb,A. and Darst,S.A. (2000) A structural model of transcription elongation. *Science*, **289**, 619–625.
12. Zhang,G., Campbell,E.A., Minakhin,L., Richter,C., Severinov,K. and Darst,S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, **98**, 811–824.
13. Murakami,K.S., Masuda,S., Campbell,E.A., Muzzin,O. and Darst,S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme–DNA complex. *Science*, **296**, 1285–1290.
14. Kalosakas,G., Rasmussen,K.O. and Bishop,A.R. (2003) Charge trapping in DNA due to intrinsic vibrational hot spots. *J. Chem. Phys.*, **118**, 3731–3735.
15. Hwa,T., Marinari,E., Sneppen,K. and Tang,L.H. (2003) Localization of denaturation bubbles in random DNA sequences. *Proc. Natl Acad. Sci. USA*, **100**, 4411–4416.
16. 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.
17. Lilley,D.M. (1988) DNA opens up–supercoiling and heavy breathing. *Trends Genet.*, **4**, 111–114.
18. Banerjee,A. and Sobell,H.M. (1983) Presence of nonlinear excitations in DNA structure and their relationship to DNA premelting and to drug intercalation. *J. Biomol. Struct. Dyn.*, **1**, 253–262.
19. Prohofsky,E.W. (1988) Solitons hiding in DNA and their possible significance in RNA transcription. *Phys. Rev. A*, **38**, 1538–1541.
20. Peyrard,M. and Bishop,A.R. (1989) Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, **62**, 2755–2758.
21. Dauxois,T., Peyrard,M. and Bishop,A.R. (1993) Dynamics and thermodynamics of a nonlinear model for DNA denaturation. *Phys. Rev. E. Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **47**, 684–695.
22. Usheva,A. and Shenk,T. (1994) TATA-binding protein-independent initiation: YY1, TFIIB and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell*, **76**, 1115–1121.
23. Malik,S., Hisatake,K., Sumimoto,H., Horikoshi,M. and Roeder,R.G. (1991) Sequence of general transcription factor TFIIB and relationships to other initiation factors. *Proc. Natl Acad. Sci. USA*, **88**, 9553–9557.
24. Benham,C.J. (1990) Theoretical analysis of heteropolymeric transitions in superhelical DNA molecules of specified sequence. *J. Chem. Phys.*, **92**, 6294–6305.
25. 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. Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **58**, 3585–3588.
26. Wiegand,R.C., Godson,G.N. and Radding,C.M. (1975) Specificity of the S1 nuclease from *Aspergillus oryzae*. *J. Biol. Chem.*, **250**, 8848–8855.
27. Berk,A.J. (1986) Adenovirus promoters and E1A transactivation. *Annu. Rev. Genet.*, **20**, 45–79.
28. Shenk,T. (1996) In Fields,B., Knipe,D. and Howley,P. (eds) *Fields' Virology,* 3rd edn. Lippincott-Raven Publishers, Philadelphia, PA, Vol. 2, pp. 2111–2148.
29. Pastor,N., Pardo,L. and Weinstein,H. (1997) Does TATA matter? A structural exploration of the selectivity determinants in its complexes with TATA box-binding protein. *Biophys. J.*, **73**, 640–652.
30. Flatters,D. and Lavery,R. (1998) Sequence-dependent dynamics of TATA-Box binding sites. *Biophys. J.*, **75**, 372–381.
31. Patel,D.J., Ikuta,S., Kozlowski,S. and Itakura,K. (1983) Sequence dependence of hydrogen exchange kinetics in DNA duplexes at the individual base pair level in solution. *Proc. Natl Acad. Sci. USA*, **80**, 2184–2188.
32. Pastor,N., Weinstein,H., Jamison,E. and Brenowitz,M. (2000) A detailed interpretation of OH radical footprints in a TBP–DNA complex reveals the role of dynamics in the mechanism of sequence-specific binding. *J. Mol. Biol.*, **304**, 55–68.
33. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
34. Wu,J., Parkhurst,K.M., Powell,R.M., Brenowitz,M. and Parkhurst,L.J. (2001) DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent. *J. Biol. Chem.*, **276**, 14614–14622.
35. Tratschin,J.D., Miller,I.L. and Carter,B.J. (1984) Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function. *J. Virol.*, **51**, 611–619.
36. Labow,M.A., Hermonat,P.L. and Berns,K.I. (1986) Positive and negative autoregulation of the adeno-associated virus type 2 genome. *J. Virol.*, **60**, 251–258.
37. Houbaviy,H.B., Usheva,A., Shenk,T. and Burley,S.K. (1996) Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl Acad. Sci. USA*, **93**, 13577–13582.
38. 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.

39. Hyde-DeRuyscher,R.P., Jennings,E. and Shenk,T. (1995) DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res.*, **23**, 4457–4465.
40. Coleman,R.A. and Pugh,B.F. (1995) Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. *J. Biol. Chem.*, **270**, 13850–13859.
41. Giudice,E. and Lavery,R. (2003) Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc.*, **125**, 4998–4999.
42. Hoopes,B.C., LeBlanc,J.F. and Hawley,D.K. (1998) Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J. Mol. Biol.*, **277**, 1015–1031.
43. Frauenfelder,H., McMahon,B.H., Austin,R.H., Chu,K. and Groves,J.T. (2001) The role of structure, energy landscape, dynamics and allostery in the enzymatic function of myoglobin. *Proc. Natl Acad. Sci. USA*, **98**, 2370–2374.