Comment on "Can One Predict DNA Transcription Start Sites by Studying Bubbles?"

In a recent Letter, van Erp *et al.* [1] have used the Peyrard-Bishop-Dauxois (PBD) model to study the following question: "Can one predict DNA transcription start sites by studying bubbles?" The purpose of this Comment is to point out that there are several key ingredients missing in this model which are essential for such a prediction and that alternative models exist which do allow one to make such predictions successfully.

First, the PBD model looks at spontaneous bubble formation as a strictly thermal process, in the absence of any superhelical stress. But *in vivo*, the bubble formation required for transcriptional initiation is stringently regulated. In part this involves modulating the superhelical stresses that act to destabilize the duplex. These stresses are crucial to the processes by which bubbles actually form during transcriptional initiation. Without them, in normal cellular conditions the chances of any sustained bubble formation is negligible. Both enzymatic and transcriptional events are involved in generating superhelical stresses, which can be strong enough to induce stable opening at susceptible sites [2]. While the sites involved are well characterized by the thermodynamic properties that underlie the PBD model, the interactions among them that determine which open under stresses are not.

An example [3] that illustrates clearly why any thermal model is inadequate is shown in Fig. 1, where the probability of bubble formation is plotted for a chicken histone gene at two levels of superhelical stress. At the lesser stress level of $\Delta Lk = -27$, a bubble forms at position 440 in the promoter region of the gene. But at a more extreme stress, this position reanneals, coupled by the stress to opening at position 4400. This shows that bubble formation driven by stresses involves complex interactions that can give nonmonotonic transitions, something that does not occur in strictly thermally driven processes.

Second, the base-pairing energies used in the PBD model depend only on the GC content. This is known not to be a complete representation of these energies. The stability of base pairing derives largely from base stacking and is not due just to hydrogen bonding. Thus, the free energy of base pairing is in large part entropic (hydrophobic) in character and depends not only on the pair under consideration but also on the identities of its nearest neighbors. It is possible to associate a free energy with each pair of stacked base pairs, as these energies have been experimentally measured [4].

Third, given the amount of available experimental knowledge of this system, it is adequately predictive to treat the base-pair separation as an Ising variable rather than a continuous variable, y, with a Morse potential as is done in the PBD model. The Ising variable can represent whether the base pair is open or closed and allows one to study long sequences. Improvements in the model are more likely to arise from considering the states of the DNA when it is not double stranded than from refining a continuous potential, as in the PBD model.

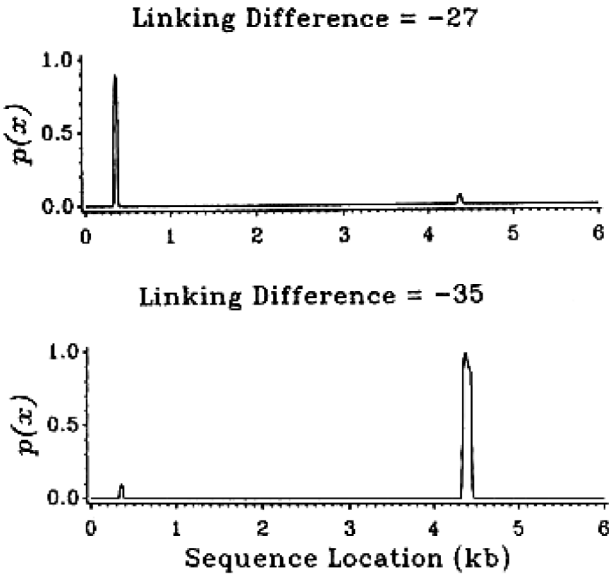

FIG. 1. Opening probability, for a chicken histone gene, vs sequence location at two levels of superhelical stress [3]. Reprinted with permission.

A model incorporating these features has been developed by one of us [5]. This approach has been used to show that sites where bubble formation is driven by stresses are closely associated with several types of genomic regulatory regions, including promoters [6] and replication origins [7].

Craig J. Benham[1] and Rajiv R. P. Singh[2]
[1]Genome Center and Department of Mathematics
University of California
Davis, California 95616, USA
[2]Department of Physics
University of California
Davis, California 95616, USA

Received 27 December 2005; published 2 August 2006
DOI: 10.1103/PhysRevLett.97.059801
PACS numbers: 87.15.Aa, 05.10.−a, 87.15.He

[1] T. S. van Erp *et al.*, Phys. Rev. Lett. **95**, 218104 (2005).
[2] F. Kouzine *et al.*, Nat. Struct. Biol. **11**, 1092 (2004).
[3] C. J. Benham, J. Mol. Biol. **255**, 425 (1996).
[4] G. Steger, Nucleic Acids Res. **22**, 2760 (1994).
[5] R. M. Fye and C. J. Benham, Phys. Rev. E **59**, 3408 (1999).
[6] H. Wang *et al.*, Genome Res. **14**, 1575 (2004).
[7] P. Ak and C. J. Benham, PLOS Comput. Biol. **1**, e7 (2005).

