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

Recently, van Erp *et al.* published a Letter [1] presenting conclusions which contradict earlier works by us [2,3]. We believe this criticism to be misguided—indeed the authors' conclusions largely substantiate the original discovery. This needs to be clarified so that the broader community is not misled.

In our earlier work we provided experimental and theoretical evidence that functional sites for transcription can coincide with thermally induced openings of double-stranded DNA. The comments by van Erp *et al.* regarding the veracity of our experiments are without foundation. The theoretical basis for these studies was provided by the simple Peyrard-Bishop-Dauxois (PBD) model. Using parameters already established in the literature, we performed Langevin simulations on the PBD model for a few specific viral DNA sequences. The object of our simulations was to establish whether there were certain regions in these highly heterogeneous real sequences that were more prone to sustain large thermally induced openings ("bubbles") of the double-stranded molecule. Our simulations indicated that there indeed were such regions, and we experimentally verified their existence using the S1 nuclease technique. Based on these combined theoretical and experimental findings we made the *observation* that in these viral sequences the regions sustaining the large bubbles coincided with known binding sites active in transcription. We concluded that it might be possible more generally to identify DNA functionally active sites, including transcription initiation sites, by studying the thermal fluctuations (particularly large amplitude coherent openings) of the double strand. It is very important to note that in fact these observations and speculations could have been made entirely based on experimental evidence. We were, however, fortunate to also possess a model (PBD) which sufficiently contained essential entropic ingredients to accurately guide the location of openings.

The claim of van Erp *et al.* is that our simulation technique was inadequate and therefore the entire work is flawed. As in our own independent elaborations [4] of our original discovery, van Erp *et al.* assume that thermodynamic equilibrium averages are sufficient and perform explicit integrations of the integrals involved in the partition function for the above model. It is therefore correct that their and our [4] results with respect to obtaining thermodynamic averages are more accurate than our initial results. However, the two crucial aspects of our initial findings are confirmed by these studies of thermodynamical equilibrium properties:

(1) In all three of the viral sequences examined by van Erp *et al.*, they find the regions which sustain large bubbles to be precisely those we originally identified [2,3], and these regions indeed include the sites active during transcription.

(2) van Erp *et al.* find that the two base pair mutation of the AAVP5 promoter causes a significant suppression of large thermal fluctuations at the former transcription sites, exactly as we reported earlier [2,3].

We emphasize again that these observations were strongly supported by experiments (notably missing in the work of van Erp *et al.*). It is true that our subsequent studies [4] have found that the relevant quantities for function is likely to be the probability of bubbles of *specific* sizes—presumably associated with physical dimensions of, e.g., transcription machinery—but this does not dilute our primary discovery.

van Erp *et al.* make one valid point in regard to our earlier work: We published results on a "control" sequence, which consisted of a nonpromoter containing a similar number of base pairs as the promoter sequences. The results we showed for this case indicated that no bubbles occurred in this sequence. More accurate Langevin results do show the occurrence of bubbles in this sequence, as noted by van Erp *et al.* This may indicate that a human coding gene was a poor choice for the control, and that our scenario is best suited to viral sequences or, that the specific bubble sizes are key. In any case this does not affect the validity of our original results for the active promoters.

In conclusion, the results of van Erp *et al.* in essence confirm our initial discovery: the PBD model is indeed an accurate guide to the location of experimentally identified active openings. Clearly this simple model must be further augmented to describe either fully realistic dynamics, or how biological machinery (such as RNA polymerase) engages these active regions. We are working to implement these augmentations.

C. H. Choi,[1] A. Usheva,[1] G. Kalosakas,[2]
K. Ø. Rasmussen,[3] and A. R. Bishop[3]
[1]Beth Israel Deaconess Medical Center
and Harvard Medical School
99 Brookline Avenue
Boston, Massachusetts 02215, USA
[2]Max Planck Institute for the Physics of Complex Systems
Nöthnitzer Strasse 38
D-01187 Dresden, Germany
[3]Theoretical Division and Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, New Mexico 87545, USA

Received 9 December 2005; published 14 June 2006
DOI: 10.1103/PhysRevLett.96.239801
PACS numbers: 87.15.He, 87.14.Gg, 87.15.Aa, 87.15.Ya

[1] T. S. van Erp *et al.*, Phys. Rev. Lett. **95**, 218104 (2005).
[2] C. H. Choi *et al.*, Nucleic Acids Res. **32**, 1584 (2004).
[3] G. Kalosakas, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, Europhys. Lett. **68**, 127 (2004).
[4] Z. Rapti, A. Smerzi, K. Ø. Rasmussen, and A. R. Bishop, C. H. Choi, and A. Usheva, Europhys. Lett. **74**, 540 (2006).

