This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Comment on "SCS: Signal, Context, and Structure Features for Genome-Wide Human Promoter Recognition"

Denise Fagundes-Lima and Gerald Weber

Abstract—We comment on the flexibility profiles calculated by Zeng et al., and show that these profiles do not represent the local flexibility of the DNA molecule. If one takes into account the physics of elasticity, the averaged flexibility profile show an additional peak which is missed in the original calculation. We show that it is not possible to calculate the flexibility of a 6-mer using tetranucleotide elastic constants, the shortest sequence is a 7-mer. For 6-mers, dinucleotide or trinucleotide parameters are needed. We present calculations for dinucleotide flexibility parameters and show that the same additional peak is present for both 7-mers and 6-mers.

Index Terms—Promoter analysis, DNA elasticity, flexibility profiles

Zeng et al., in their work on promoter recognition [1], introduce an index which uses tetranucleotide flexibility parameters obtained from Ref. 2. These flexibility parameter have the physical dimension of an elastic constant per mol, see Eq. 1 in Ref. 2. As described in their Eqs. (3–5), Zeng et al. use a simple consecutive summation of these elastic constants. Specifically, they use three consecutive tetranucleotide parameters to calculate a 6-mer index f

$$f_i = t_{i,i+3} + t_{i+1,i+4} + t_{i+2,i+5}, \tag{1}$$

where i is the starting position of the 6-mer in the genomic sequence.

To understand how to obtain the elastic constant of a longer sequence from shorter segments, one has to imagine a DNA molecule as composed of a certain number of coupled springs. Since these springs are chained tail-to-head to each other, it is possible to calculate an equivalent elastic constant which represents the longer molecule. For a 7-mer we would use two tetramers with one overlapping position which is where the two tetramers are chained together. It is therefore not possible to obtain a 6-mer elastic constant from tetramers. Furthermore, it the equivalent elastic constant of a 7-mer is obtained from the inverse summation of the elastic constants of the two tetramers. A direct summation as proposed in Eq. (1) would be representative of 3 tetramers in parallel which does not represent the physical configuration of DNA elasticity. As a result, the index f in Eq. (1) does not represent the elastic property of an 6-mer but has become essentially an arbitrary index.

Nevertheless, the index f is built from elastic constants. It is therefore legitimate to ask to which extent does this index compare to a flexibility profile? To answer this question we calculated the equivalent elastic constants for 7-mers using the same human promoter sequences from DBTSS (version 5.2.0) [3] and tetranucleotide parameters from Ref. 1.

For a 7-mer the resulting equivalent elastic constant $t_{1,7}$ is calculated from tetranucleotide elastic constants $t_{i,i+3}$,

$$\frac{1}{t_{1,7}} = \frac{1}{t_{1,4}} + \frac{1}{t_{4,7}}, \tag{2}$$

where the subscripts i, j represent the start and end position of the segment. Eq. 2 results from a straightforward application of Hookes law and is easy to understand intuitively. Given two coupled springs, say one soft and the other rigid, the force exerted will deform much more the softer spring. Overall, the chained springs are easier to deform than each individually. Therefore, the resulting equivalent elastic constant is dominated by the softer part of the elastic constant. Using the same example as in Eqs. (4–5) of Ref. 1, the equivalent elastic constant of a 7-mer becomes

$$\frac{1}{t_{\mathrm{TATAAAA}}} = \frac{1}{t_{\mathrm{TATA}}} + \frac{1}{t_{\mathrm{AAAA}}}. \tag{3}$$

This highlights yet another problematic aspect of the parameter f which is that the central part of the molecule becomes over-represented.

Unavoidably, using either Eq. (1) or Eq. (2) will result in different profiles. In Fig. 1a we show the profiles recalculated from Ref. 1. Using Eq. (2) we observe additional peak in Fig. 1b, close to transcription starting position, which is missed altogether when using Eq. (1). Interestingly, the region around -28 retains its rigid character even when the flexibility is calculated according

- *D. Fagundes-Lima is with the Department of Biological Sciences, Federal University of Ouro Preto, Ouro Preto-MG, Brazil, defalima@gmail.com*
- *G. Weber is with the Department of Physics, Federal University of Minas Gerais, Belo Horizonte-MG, Brazil.*

Digital Object Indentifier 10.1109/TCBB.2011.130

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

to Eq. (2). This result is surprising and non-trivial since Eq. (2) generally favours softer elastic constants. It also confirms the interpretation from Zeng *et al.* that this region is notably rigid when tetranucleotide parameters from Ref. 2 are used.

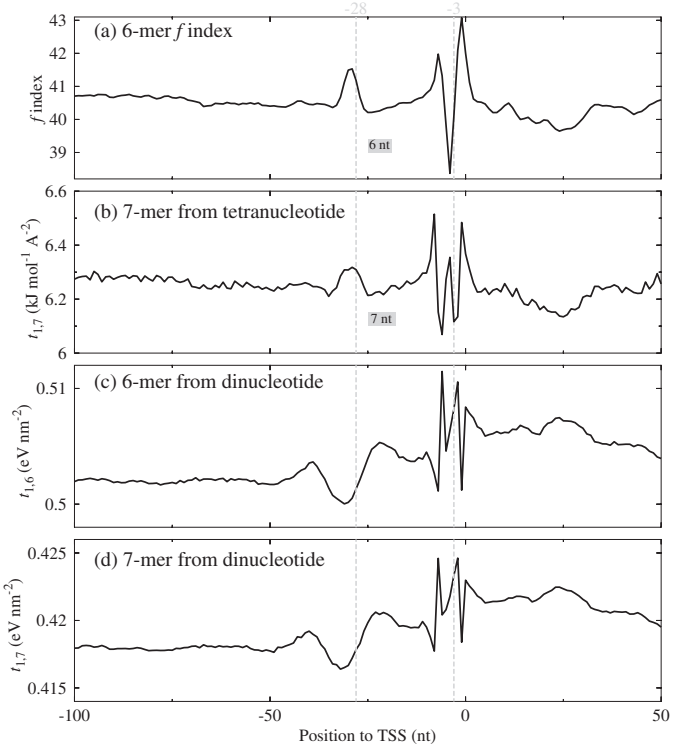

Fig. 1. Promoter profiles for the (a) 6-mer f-index, (b) 7-mer equivalent elastic constant, and elastic constants calculated using dinucleotide parameters for (c) 6-mers and (d) 7-mer. The gray boxes show the width of a 6-mer and a 7-mer sliding window.

The comparison of Figs. 1a and 1b raises two further questions. The first question is whether the additional peak in Fig. 1b does result from Eq. (2) or from a longer sliding window? The next question is how should one proceed to calculate profiles for 6-mers? Starting with the second question, one way to obtain the elastic properties of 6-mers would be to use nucleotide parameters of dimers or trimers. For instance, using dinucleotide flexibility parameters one can generalise Eq. (2) for N-mers

$$\frac{1}{t_{1,N}} = \sum_{i=1,\ldots,N-1} \frac{1}{t_{i,i+1}}. \qquad (4)$$

Figure 1c and 1d shows the flexibility profiles calculated using dinucleotide elastic constants from Ref. 4. Both 6-mer and 7-mer flexibility profile show nearly identical results. Therefore, it seems reasonable to assume that the missed peak of Fig. 1a results from the way the f index is calculated and not from a shorter sliding window.

We take this opportunity to comment on the use of flexibility parameters from different methods and experiments [2], [4], as these provide some interesting insights into the flexibility properties of promoters. Packer *et al.* [2] used data from X-ray diffraction measurements to obtain their elastic constants. This type of measurements probes essentially the static configuration of the DNA molecule and from Fig. 1b one is lead to conclude that the region around -28 should be largely rigid. On the other hand, the parameters from Ref. 4 result from melting temperatures and as such probe the dynamics of the DNA molecule. In this case Figs. 1c,d indicate an exceptionally soft region around -28, in contrast to Fig. 1b.

In conclusion, we show that the f-index is not the appropriate representation of the elasticity of the DNA molecule. However, this does not imply in its inadequacy for promoter recognition. Zeng et al. [1] showed that the f-index is useful for promoter recognition and this remains unchanged. Our finding only concerns the interpretation of these results in terms of DNA elasticity, and draws attention to the problems which arise when comparing f-profiles with flexibility profiles.

ACKNOWLEDGEMENTS

This work was funded by CNPq, Fapemig and National Institute of Science and Technology for Complex Systems.

REFERENCES

[1] J. Zeng, X.-Y. Zhao, X.-Q. Cao, and H. Yan, "SCS: Signal, context, and structure features for genome-wide human promoter recognition," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, pp. 550–562, 2010.

[2] M. J. Packer, M. P. Dauncey, and C. A. Hunter, "Sequence-dependent DNA structure: tetranucleotide conformational maps," *J. Mol. Biol.*, vol. 295, no. 1, pp. 85–103, 2000.

[3] R. Yamashita, Y. Suzuki, H. Wakaguri, K. Tsuritani, K. Nakai, and S. Sugano, "DBTSS: database of human transcription start sites, progress report 2006," *Nucleic Acids Research*, vol. 34, no. suppl 1, p. D86, 2006.

[4] G. Weber, J. W. Essex, and C. Neylon, "Probing the microscopic flexibility of DNA from melting temperatures," *Nature Physics*, vol. 5, pp. 769–773, 2009.

Denise Fagundes Lima received the BSc degree in biological sciences from the Federal University of Ouro Preto, Brazil, in 2009. She is now completing her MSc in Biotechnology also at the Federal University of Ouro Preto. Her research interests include bioinformatics and the biology of promoters and microRNA.

Gerald Weber received the PhD degree in physics from the State University of Campinas, Brazil, in 1990. He worked with theoretical semiconductor physics until 2003 when he changed his research interest to the physics of DNA and bioinformatics when at the School of Chemistry of the University of Southampton, UK. Since 2009 he is a lecturer at the Department of Physics of the Federal University of Minas Gerais, Brazil. His current research interest are in thermodynamical properties of DNA and RNA and its application to problems in computational biology.