

# Soft Power of Nonconsensus Protein-DNA Binding

Vladimir B. Teif<sup>1,\*</sup>

<sup>1</sup>School of Life Sciences, University of Essex, Colchester, United Kingdom

In this issue of *Biophysical Journal*, Goldshtein et al. (1) show that if gene promoters are extended with DNA sequences containing repeating nucleotide patterns without specific protein-binding motifs, it is possible to predict the resulting changes in gene expression from so-called nonconsensus protein-DNA binding. The authors found that during embryonic stem cell (ESC) differentiation, transcription factor (TF) preferences for such simple nucleotide repeats undergo distinct changes. This suggests an intriguing possibility that nonconsensus binding may help direct TFs to different subclasses of binding sites in different cell types.

DNA-protein binding has been studied in great detail for about a century. Chromatin proteins usually have positively charged domains that are naturally attracted to the negatively charged DNA, whatever the nucleotide sequence is. Such binding is traditionally called nonspecific (2). However, most biologically interesting processes depend on the DNA sequence—this is when sequence-specific binding comes into play. In many classical examples of DNA-protein binding, a TF recognizes a single stretch of nucleotides on the DNA (3), the so-called consensus binding motif and its variations. Usually, the strength of such

binding is several orders of magnitude higher than that of nonspecific binding, which led to the concept of discrete TF binding sites—a limited number of small genomic regions where a given TF can bind (as opposed to the rest of the genome, where TF binding can still happen but can be neglected because of its weakness). The concept of discrete TF binding sites has been very useful in predicting combinatorial, cooperative TF binding for several decades. Not all TFs appeared to recognize a single motif, but the concept of discrete binding sites could still hold by allowing several motifs for a single TF. The discrete binding site concept can be further extended to take into account new “letters” of the DNA alphabet arising because of naturally occurring chemical modifications, such as methylation. However, in recent years, with the arrival of advanced computational methods, such as deep learning on one hand and affordable high-throughput experiments on the other, it is becoming increasingly clear that many important TF-DNA binding events occur in the intermediate regime between nonspecific and discrete site binding (4).

One possibility for describing such an intermediate binding regime is to widen the definition of the classical concept of consensus motifs to include, in addition to motifs responsible for unique DNA structures recognized by a single protein, new motifs responsible for several generic classes of local shape of the DNA double helix (such as widening of the narrow

groove, bending, etc.) (5,6). Another possibility is to characterize protein-DNA binding beyond sequence motifs, searching for repeating nucleotide patterns—this is the approach that Goldshtein et al. (1) took (Fig. 1).

Investigations of the role of genomic nucleotide periodicities in DNA-protein binding started about four decades ago but were mainly in the context of nucleosome positioning (7). Nucleosomes cover most of the genome, so the concepts of a discrete binding site and a consensus motif naturally do not apply in this case, whereas the concept of nucleotide (or dinucleotide) oscillations appears quite handy. Nucleosome positioning is known to affect TF binding through competitive and cooperative interactions (8), but the direct influence of DNA nucleotide periodicities on TF binding considered by Goldshtein et al. (1) is a separate important effect.

In a series of recent publications, Lukatsky and colleagues (9) argued that stronger-than-random interaction of a DNA-binding protein with a genomic region containing simple nucleotide repeats has an entropic nature; in this case, it is statistically more probable for a protein to land on a region that contains its “favorite” repeat element (Fig. 1 D). They have performed in vitro experiments demonstrating that, for natural genomic sequences, the strength of this effect is quite significant and comparable to that of mutations in the specific TF motif (9). In their most recent work Goldshtein et al. (1) applied this

Submitted February 25, 2020, and accepted for publication February 26, 2020.

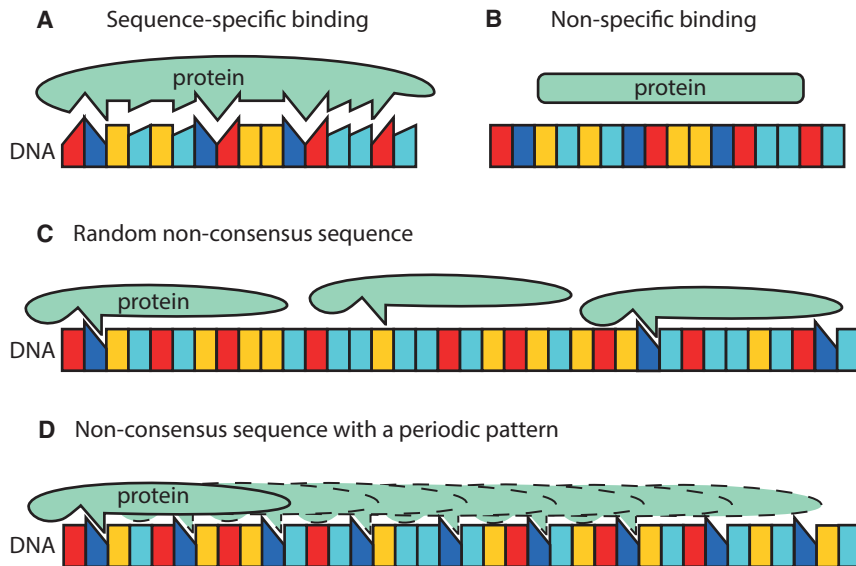
\*Correspondence: [vteif@essex.ac.uk](mailto:vteif@essex.ac.uk)

Editor: Wilma Olson.

<https://doi.org/10.1016/j.bpj.2020.02.026>

© 2020 Biophysical Society.





**FIGURE 1** Protein-DNA binding is not limited to sequence specific (A) and nonspecific (B). It can also be characterized by an intermediate regime of nonconsensus binding, which is nonspecific for random DNA sequences (C) but has an increased binding strength for DNA regions enriched with certain repeated nucleotide patterns (D). The DNA lattice units shown in different colors may correspond to individual basepairs or larger regions. Different geometric shapes of such units used on the figure do not imply that such shapes are actually visible in the DNA structure—these may also correspond to alterations of the DNA double helix stability or other slight perturbations of the energy landscape. To see this figure in color, go online.

approach to characterize TF-DNA binding in ESCs. For example, the authors showed that one of the key regulators of ESC development, c-Myc, possesses statistical preference for repetitive patterns of the type [CNNC] and [GNNG], where N stands for any nucleotide type. This computational prediction was verified in a plasmid reporter assay, introducing such repetitive patterns surrounding the consensus c-Myc-binding sites at the promoter of the reporter gene. As predicted, this resulted in higher gene expression in comparison with the case in which flanking regions around c-Myc sites were composed of random DNA sequences.

The question about the physical nature of such nonconsensus TF-DNA binding is still open. One possibility is

that nonconsensus DNA sequence repeatedly modulates the local stability of the DNA double helix. Goldshtein et al. (1) performed NMR measurements showing that DNA sequences with identical GC content but different DNA-repeat-symmetry types can indeed lead to different local DNA stabilities. Another possibility is that the nonconsensus binding effects are due to slight changes of the local DNA shape, as in the models with DNA-shape motifs (5,6), but in this case because of nucleotide oscillations in the absence of well-defined motifs. In a recent study that investigated 100 million random promoters, the effects on *cis*-regulatory logic associated with nucleotide changes outside TF binding motifs were mostly interpreted through the changes of DNA accessibility (10),

but in light of the works mentioned above, this can be also explained by direct effects on nonconsensus TF binding. Since genomes contain multitudes of simple repeats, such effects may have important roles in guiding differential TF binding during cell transitions, executing “soft power” on gene regulation beyond consensus motifs.

## REFERENCES

1. Goldshtein, M., M. Mellul, ..., D. B. Lukatsky. 2020. Transcription factor binding in embryonic stem cells is constrained by DNA sequence repeat symmetry. *Biophys. J.* 118:2015–2026.
2. von Hippel, P. H., A. Revzin, ..., A. C. Wang. 1974. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc. Natl. Acad. Sci. USA.* 71:4808–4812.
3. Ptashne, M. 1967. Specific binding of the lambda phage repressor to lambda DNA. *Nature.* 214:232–234.
4. Inukai, S., K. H. Kock, and M. L. Bulyk. 2017. Transcription factor-DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.* 43:110–119.
5. Samee, M. A. H., B. G. Bruneau, and K. S. Pollard. 2019. A de novo shape motif discovery algorithm reveals preferences of transcription factors for DNA shape beyond sequence motifs. *Cell Syst.* 8:27–42.e6.
6. Zhou, T., N. Shen, ..., R. Rohs. 2015. Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. USA.* 112:4654–4659.
7. Trifonov, E. N., and J. L. Sussman. 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA.* 77:3816–3820.
8. Teif, V. B., F. Erdel, ..., K. Rippe. 2013. Taking into account nucleosomes for predicting gene expression. *Methods.* 62:26–38.
9. Afek, A., J. L. Schipper, ..., D. B. Lukatsky. 2014. Protein-DNA binding in the absence of specific base-pair recognition. *Proc. Natl. Acad. Sci. USA.* 111:17140–17145.
10. de Boer, C. G., E. D. Vaishnav, ..., A. Regev. 2020. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* 38:56–65.