New Insights into Protein-DNA Binding from Simulations of DNA Breathing Dynamics

Boian S. Alexandrov, T-1;
Amy L. Bauer, XTD-4;
Alan R. Bishop, PADSTE; William S. Hlavacek,
Fangping Mu, T-6
Elizabeth Hong-Geller, B-10;
Kristy Nowak-Lovato, D-3;
Kim Ø. Rasmussen, T-5;
Ludmil B. Alexandrov, Wellcome Trust Sanger Institute
Anny Usheva, Harvard Medical School

Understanding the mechanisms governing the affinities of transcription factors (TF) for their DNA binding sites is of tremendous importance for the regulation of gene expression in the cell. The DNA physicochemical properties that are most relevant to the TF-DNA binding have yet to be fully understood. Recently, we investigated the role of DNA breathing dynamics as a mechanical property of DNA regulating TF-DNA binding, and specifically the relationship between the DNA local propensity for breathing and binding of two well-known TFs: (1) the human TF YY1, and (2) the nucleoid-associated protein Fis in *Escherichia coli*. Our results demonstrate that suppressing or enhancing the DNA propensity for local transient opening at known binding sites, for example, by introducing a few base pairs (bp) substitutions or nucleotide modifications such as methylation (designed by computer simulations to change DNA breathing without changing the direct points-of-contact) can drastically affect TF binding.

Cellular proteins are TFs that regulate gene expression, and thereby cellular activities and fate, by binding to specific DNA segments positioned at different regulatory locations in the genome. Locations and specificities of a TF's binding site are determined by a complex interplay of specific direct points-of-contact (direct recognition) and by a set of physicochemical properties of DNA, such as local shape (indirect recognition). Despite the important role of TFs in myriad cellular processes and diseases, the physical mechanisms underpinning the strength of binding (affinity) and recognition (specificity) are still far from being completely understood. Recently developed high-throughput techniques, such as chromatin immunoprecipitation coupled with microarrays, next generation sequencing, or exonucleases and protein binding microarrays (PBM), have enabled large-scale profiling and characterization of binding sites and revealed a complex landscape of DNA binding preferences [1,2]. Only a fraction of the observed TF binding sites can be explained by direct (or protein-partner assisted) binding or captured with a single consensus sequence or position weight matrix. Even the characterization of cell-specific chromatin structure and leveraging of sophisticated motif recognition protocols still left surprising exceptions in which the binding, or the lack of it, is hard to explain [2]. A possible reason is that the bioinformatic models utilized to explain and predict TF binding sites, which rely on letter-code representations of DNA sequences, neglect physicochemical features of DNA that are important for protein-DNA recognition [3]. Here we highlight two recent studies showing that a specific physical property of DNA, namely the propensity for transient opening and re-closing of the double helix, resulting from inherent thermal fluctuations, also known as "DNA breathing" or "DNA bubbles," is associated with binding affinity, at least in the case of Fis and YY1, both of which are well-studied TFs. Using a DNA mesoscopic nonlinear model [4] that can be used for rational design of local DNA local breathing [5], we simulated breathing dynamics of known or engineered Fis and YY1 binding sites [6,7] and found a strong association between the propensity for breathing (at the binding sites) and YY1 and Fis binding affinity (at the same sites) (Fig 1, A).

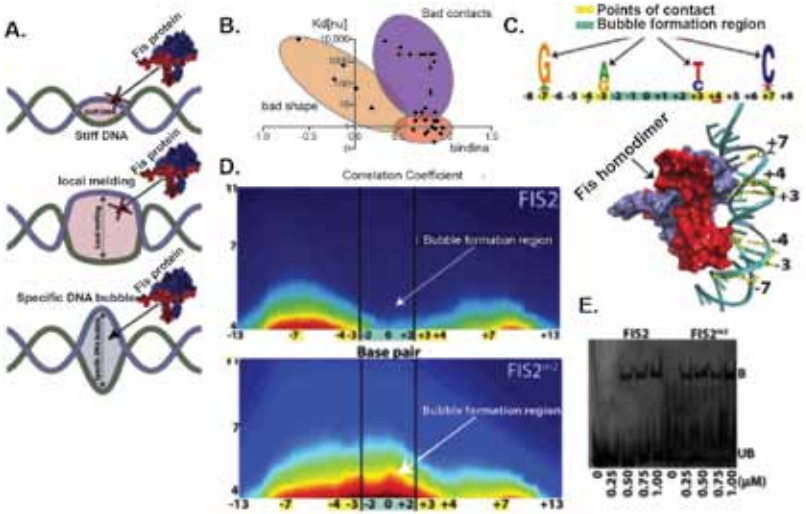

Fig. 1. A: Schematic representation of DNA bubble formation in Fis-DNA binding–(i) Weak Fis-DNA binding to stiff sequences; (ii) Weak Fis-DNA binding to significant destabilization of the double helix; and (iii) Strong Fis-DNA binding in the presence of a characteristic breathing profile. B: Schematic affinity shape-correlation diagram of in vitro sequences. Each point represents an oligomer characterized by 1) specific direct points-of-contact, 2) correlation with the characteristic opening profile, and 3) dissociation constant. C: The Fis logo and crystal structure example of Fis-DNA binding complex [8]. D: Langevin dynamics simulations demonstrating enhanced FIS2m2 local DNA breathing dynamics . FIS2m2 was designed by introducing two O6-methylguanine in the bubble formation region of the FIS2 sequence, without changing the points-of-contacts [7]. The color map represents the probability of bubble opening; high (red) and low (blue) probability of opening. The length of the transient bubbles, in base pairs [bp], is shown along the vertical axis. The horizontal axis depicts bp position; the bubble formation region is in blue while the points-of-contact are in yellow. E: EMSA demonstrating the increase in complex formation of FIS2m2 versus FIS2. Adapted from [7].

Fis: Our simulation data and available *in vitro* Fis binding data indicate a strong correlation between DNA breathing dynamics and Fis binding (Fig. 1, B). The Fis direct points-of-contact are predominantly on one side of the double-stranded DNA binding segment [8] (Fig 1, C). It is natural to expect that the binding of a protein with direct points-of-contact predominantly on one of the two DNA strands, such as Fis, can be promoted by transient openings of the DNA at the TF binding site. Leveraging simulations of the breathing of known Fis *in vitro* binding sites, we defined an average DNA breathing profile that is characteristic of a strong

For more information contact Boian Alexandrov at boian@lanl.gov.

Funding Acknowledgments
National Institutes of Health; American Recovery and Reinvestment Act; Harvard Univeristy William F. Milton Award;
LANL, Laboratory Directed Research and Development Program

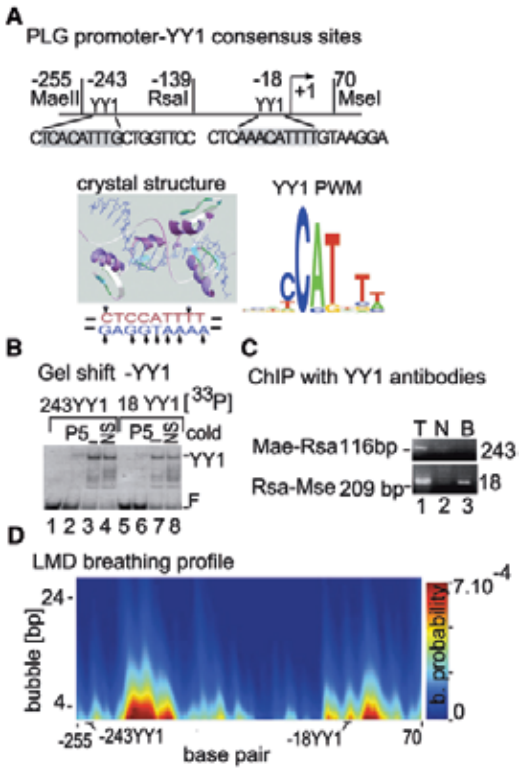

Fig. 2. A: YY1 binding sites locations relative to the transcription start site (TSS) (+1) of the human PLG promoter. The crystal structure of the YY1-P5 DNA complex [9] and the base-specific YY1 contacts on both strands (arrows). The YY1 logo is shown on the right. B: Gel shift assays demonstrate recombinant YY1 binding to 24 bp long synthetic oligonucleotides containing the two sequences: 243YY1, respectively, 18YY1 (0.1 nM), as indicated above the plots. C: ssChIP assay is used to verify genomic YY1 binding at the identified consensus sequences. After sonication and restriction enzymes digestion the YY1 antibody-captured promoter fragments are amplified by PCR with fragment-specific primers; line 1–total DNA before antibody selection (T); line 2–pulled-down DNA with control antibody (N); line 3–pulled-down with YY1 antibody (B). The identity of the PCR-amplified fragments is shown at the left: Hae-Rsa contains the 243YY1; the 18YY1 site is located in the Rsa-Mse restriction fragment. D: LMDs simulations demonstrating local DNA breathing dynamics in a 305-bp-long human PLG promoter fragment. The notations are as in Fig. 1. Adapted from [5].

Fis binding site and demonstrated that this profile is significantly enriched among known *in vivo E. coli* Fis binding sites. Furthermore, to test our understanding of how Fis binding is influenced by DNA breathing dynamics, we numerically designed base-pair substitutions and O6-guanine-methylation modifications in DNA segments (Fig. 1, D) that are known to interact strongly or weakly with Fis. In each case we sought to make local DNA breathing dynamics either closer to or farther from the breathing profile characteristic of a strong Fis binding site without changing the direct points-of-contact. For the modified DNA segments, we found that Fis-DNA binding, as assessed by electrophoretic mobility shift assay, is changed in accordance with our expectations (Fig. 1, E). We conclude that DNA breathing dynamics, which depends on DNA mechanics, and can be modified by nucleotide modifications such as nucleotide methylation, regulates Fis binding.

The regulation of DNA mechanics by nucleotide (i.e., cytosine, guanine, or adenine) methylation that can change the local breathing propensity of a regulatory genomic segment could play an important role in a novel gene-silencing mechanism. Such a mechanism will be based not on hypo- or hyper-methylation of the promoter segment, which leads to a dense chromatin structure [10], but on the presence of a few methylated nucleotides located at specific positions that can suppress or enhance binding of various TFs via alteration of local DNA breathing.

YY1: By using site-specific chromatin immunopecipitations (ssChIP), EMSA, BIOBASE data, and our computational framework [8], we found a specific DNA breathing profile associate with YY1 binding sites in vivo. In particular, we found that genomic flanking sequence variations and single nucleotide polymorphisms (Fig 2, A) may exert long-range effects on DNA dynamics and thereby influence YY1 binding in vivo. The correlation between our simulations and ssChIP data clearly emphasize the role of the local propensity for breathing at the YY1 binding site. For example, we found that, although YY1, in vitro, strongly binds two short DNA segments (identical with two DNA segments at the human PGL promoter) that contain the YY1 consensus motif
(Fig 2, B), the same sequences tested for YY1 binding in vivo, via ssChIP, demonstrated a lack of binding for one of them (Fig 2, C). We interpreted this phenomenon as an influence of the genomic flanks on the breathing of the YY1 binding sites, (Fig. 2 D). YY1 is a Zn-finger protein that binds to the major groove without conformational changes in the DNA or the protein itself [9]. The four YY1 fingers make multiple contacts with the major groove edges of bases but most of the specific hydrogen-based contacts are restricted to one of the two DNA strands (Fig. 2, A), as in the case of Fis. Therefore, it seems natural that the requirements for enhanced transient DNA openings at the points-of-contact region predicted by simulations which account for the effects of flanking sequences, would facilitate specific YY1 binding.

[1] Badis, G. et al., *Science* **324**, 1720 (2009).

[2] Arvey, A. et al., *Genome Res* **22**, 1723 (2012).

[3] Maienschein-Cline, M. et al., *Nucleic Acids Res* doi: 101093/nar/gks771 (2012).

[4] Bishop, A.R. et al., *Entropy–Driven Conformations Controlling DNA Functions,* Verlag-Berlin-Heidelberg, Springer (2012).

[5] Alexandrov, B.S. et al., *Nucleic Acids Res* **38**(6), 1790 (2010).

[6] Alexandrov, B.S. et al., *Nucleic Acids Res* **40**(20), 10116 (2012).

[7] Nowak-Lovato, K. et al., "Binding of Nucleoid-associated Protein Fis to DNA is Regulated by DNA Breathing Dynamics," *PLOS Comp Biol*, http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1002881 (2012).

[8] Stella, S. at al., *Genes Dev* **24**, 814 (2010).

[9] Houbaviy, H.B. et al., *Proc Natl Acad Sci USA* **93**, 13577 (1996).

[10] Tate, P.H. and A.P. Bird, *Curr Opin Genet Dev* **3**(2), 226 (1993).