Biophysical Journal
New and Notable

Biophysical Society

It is in the flanks: Conformational flexibility of transcription factor binding sites

Tsu-Pei Chiu,[1] Jinsen Li,[1] Yibei Jiang,[1] and Remo Rohs[1,2,*]
[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California and [2]Departments of Chemistry, Physics & Astronomy, and Computer Science, University of Southern California, Los Angeles, California

Transcription factors (TFs) utilize a wide range of DNA binding sites in the genome to regulate gene expression. Despite much research (1,2), TF-DNA binding mechanisms are still not completely understood, in part because there is no simple readout code that directs TFs to their genomic target sites (3). Apart from the primary sequence of the target DNA and its conformational landscape, the structural dynamics of DNA could play an important role in TF recognition. More than four decades ago, Bansal and colleagues were among the first to structurally explain the conformational flexibility of the double helix and its polymorphisms. They related the flexibility to the variability of torsion angles in the phosphodiester backbone of double-stranded DNA (4).

Protein readout mechanisms of the DNA core binding site were the focus of investigation over the years (5). However, flanking regions that are not directly contacted by TFs can also play a role in the recognition process. A role of flanking regions in affecting TF binding specificity was first identified for genomic flanks of E-box target sites surrounding basic helix-loop-helix TF binding sites (6), although that previous study defined DNA shape without the presence of conformational flexibility. Intrinsic dynamics embedded in the structure of flanking regions may influence TF-DNA binding affinity. Bansal and co-workers recently related DNA structural features to in vitro-derived binding affinity for different eukaryotic TFs (7). Their study revealed that TF binding affinity correlates with DNA structural features that are influenced by flanking regions surrounding TF-binding targets in the genome (7). Nevertheless, the complete mechanisms remained unclear.

In a new study, Ghoshdastidar and Bansal investigated the effect of conformational flexibility as an additional mechanism in TF-DNA readout (8). To achieve this, they used molecular dynamics (MD) simulations as an atomistic probe to measure intricate internal motions in TF-binding targets. The authors found that, for DNA targets of the *Drosophila melanogaster* Hox heterodimer Ultrabithorax (Ubx) and its cofactor Extradenticle (Exd), the flanking regions influenced the conformational flexibility of the core binding sites. The authors ran MD simulations for Exd-Hox heterodimers in complex with DNA fragments of 20 base pairs (bp) in length. Of these 20 bp, the central 8 bp represented the core binding site, 4 bp in each flank were variable, and GC dinucleotides capped the oligonucleotides at each end (8). The different flanks selected for the Exd-Ubx binding site represented combinations of high and low binding affinity derived from high-throughput binding assays, following a previously described protocol (7). The authors compared the results for Exd-Ubx binding with another *D. melanogaster* Hox protein, Sex combs reduced (Scr), in complex with Exd with one DNA fragment where the Exd-Scr binding site was surrounded by A-tracts in both the 5′ and 3′ flanks. This sequence with A-tract-containing flanks represented a low-affinity binding site (8).

During the MD simulations, the authors observed a structural transition from the flanks into the cores. They concluded that the flanks were essential in determining the conformational flexibility of the core binding sites. Different flanking sequences led to various degrees of conformational flexibility of the cores. This mechanism might contribute to the variation in binding affinity observed for different sequences (8). This finding might also explain the influence of flanking sequences on TF binding specificity, as was observed in prior studies (6). The study (8) provides the first biophysical explanation, to our knowledge, of the "shape space" mechanism, whereby the conformational flexibility of the flanks acts as a mediator between DNA sequence and shape. Thereby

Submitted August 29, 2022, and accepted for publication September 16, 2022.

*Correspondence: rohs@usc.edu

Editor: Tamar Schlick

https://doi.org/10.1016/j.bpj.2022.09.020

© 2022 Biophysical Society.

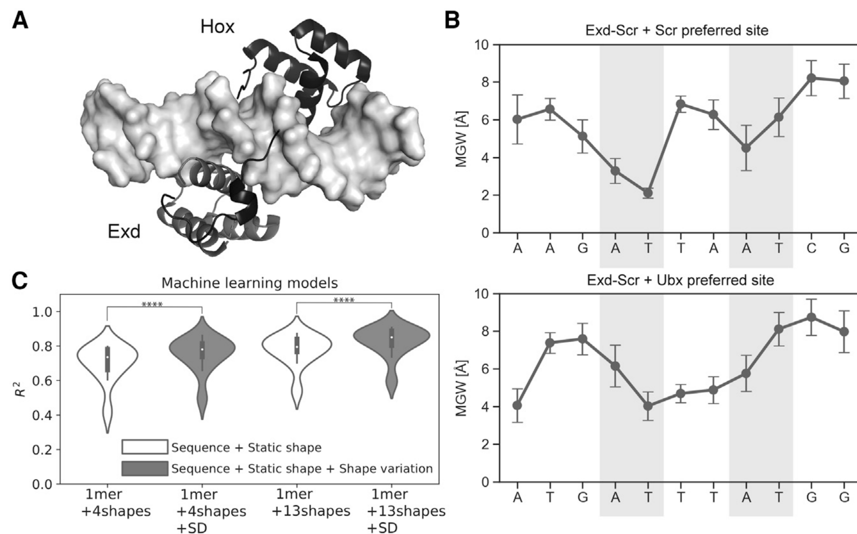

FIGURE 1 Conformational flexibility of Hox-DNA binding sites. (*A*) Overview of a co-crystal structure of an Exd-Hox complex, with Scr and the DNA target site 5′-TAAGATTAATCGG-3′ (PDB: 2R5Z). (*B*) DNA minor groove width profiles as a function of nucleotide position for Exd-Scr bound complexes in both panels. The minor groove width profiles are derived from all-atom MD trajectories, with Scr preferred binding site 5′-CTCTAAGATTAATCGGCTG-3′ (upper panel) and Ubx preferred binding site 5′-CTCTATGATTTATGGGCTG-3′ (lower panel). AMBER force fields (AMBER94 for nucleic acids and AMBER99SB for proteins) (11) were used for the MD simulations. DNA shape parameters were computed with CURVES (12) for MD snapshots every 100 ps using Trj2Shape (13). (*C*) Violin plots directly comparing R^2 values between sequence + static shape models (white) and shape fluctuation-augmented models (gray) for two different models consisting of sequence and either 4 or 13 DNA shape features (14). Models were trained by L2-regularized multiple linear regression using 10-fold cross validation on 21 datasets derived from SELEX-seq experiments for the Hox TF family (14). Statistical significance upon adding conformational flexibility to the models in form of standard deviation (SD) of DNA shape features is indicated by **** representing $p < 0.0001$.

the study introduces a previously unknown binding mechanism, in which the function of conformational flexibility is mediated by flanking regions of a TF binding site (8).

Running 500-ns MD simulations for each complex, the authors report how each flanking sequence affects the shape and flexibility of the Exd-Ubx or Exd-Scr binding site with the respective protein heterodimers bound to target sequences. The observed effect of the different flanking sequences may uncover a biophysically interesting mechanism that requires a larger scale study of additional flanking sequences. In particular, the Exd flank is the site of an additional cofactor, Homothorax (Hth), required for in vivo function of Hox proteins. The cooperative binding of the Hth-Exd-Hox trimer refines the role of the Exd flank of the Exd-Hox binding site as a spacer between the Hth and Exd cofactors with previously studied sequence and shape preferences (9). In this context, the choice of an A-tract as the only flank for the Exd-Scr binding site is limiting. A-tracts, or runs of three or more consecutive ApA, ApT, or TpT bp steps, are rigid elements due to interstrand hydrogen bonds. For the MD simulation where the Exd-Scr binding site is flanked by A-tracts, the authors report that the binding site loses its double minimum profile in minor groove width observed in the co-crystal structure of the complex and in all-atom Monte Carlo simulations of the unbound DNA target (10). This result might be due to the A-tract flanks operating as off-the-scale determinants of the binding site conformation. An important follow-up would therefore be to expand this set of MD simulations to sequences that are more likely to occur in vivo.

To answer this question, we ran all-atom MD simulations for two complexes: namely, the Exd-Scr heterodimer bound to either an Scr or Ubx preferred DNA sequence based on co-crystal structures that we previously analyzed, PDB: 2R5Y and 2R5Z (Fig. 1 *A*) (10). We reconstituted missing residues of the linker region and the conserved N-terminal arm in these structures by homology modeling. Our two 200-ns MD simulations were for the Scr preferred sequence 5′-CTCTAAGATTAATCGGCTG-3′ and the Ubx preferred sequence 5′-CTCTATGATTTATGGGCTG-3′. When we analyzed the MD simulation trajectories at the interval of 50–200 ns, we observed that the minor groove width profiles in the co-crystal structures (two minima in the Scr preferred sequence versus one minimum in the Ubx preferred sequence) were retained in the MD simulations (Fig. 1 *B*). Rather than contradicting the findings reported by the authors (8), our results instead highlight the importance and dependency on the sequence identity of the flanks and resulting conformational flexibility of the core binding site.

Based on MD simulations of the Exd-Scr heterodimer with the binding site that lost its experimentally observed minor groove width profile, the authors also report that the Scr linker residue His-12 was ejected from the minor groove (8). However, it is not clear whether this ejection led to a shape change in minor groove width profile. Histidine is usually a neutral amino acid and would therefore prefer to be solvated, which is what the authors reportedly observed. However, histidine is likely protonated in the vicinity of highly charged nucleic acids such as DNA (10). If assigned a positive charge, protonated His-12 would likely remain in the minor groove or close to the phosphodiester backbone, seeking a favorable electrostatic interaction between negatively charged DNA and the positive charge carried by His-12. Indeed, we observed in MD simulations initiated from the co-crystal structure with PDB: 2R5Z that the minor groove insertion of His-12 in its neutral state depends on the

linker's starting configuration, whereas His-12 in its protonated state remains in the minor groove throughout the MD simulation. The protonation state of His-12 affects the shape of the minor groove. When His-12 is positively charged, there is a greater tendency toward groove narrowing. This change would not alter the fact that flanking sequences affect conformational flexibility, but the balance and resulting effect on the minor groove geometry varies.

To study an unlimited number of sequences, the conformational flexibility of TF binding sites can be examined through additional methods. One such approach is the high-throughput prediction of DNA shape and consideration of fluctuations of structural features in combination with statistical machine learning. We previously introduced standard deviations of 13 different DNA shape features (14). Using experimental data for relative binding affinity of Exd-Hox heterodimers derived from SELEX-seq binding assays (15), we showed that adding conformational flexibility to static DNA shape improves the ability to predict TF binding specificity (Fig. 1 *C*) (14). The results of our machine learning study, using multiple linear regression and cross validation, agree with the current findings and conclusions of Ghoshdastidar and Bansal (8), albeit without the detailed atomic insights provided by MD simulations.

Bansal and colleagues taught us about the conformational flexibility of the double helix (4). They also contributed to our increased understanding of the functional importance of conformational flexibility in transcription (16). The work presented here (8) clearly emphasizes the importance of the conformational flexibility of TF binding sites and proposes a mechanism for how it is achieved. Whereas more studies are needed to understand how flanking regions are utilized by TFs, the authors provide evidence that the genome modulates conformational flexibility through flanking regions that are otherwise not contacted by TFs or are not part of TF core binding sites.

ACKNOWLEDGMENTS

The authors acknowledge support from the National Institutes of Health (grant R35GM130376 to R.R.) and the Human Frontier Science Program (grant RGP0021/2018 to R.R.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. 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. https://doi.org/10.1016/j.gde.2017.02.007.
2. Zeitlinger, J. 2020. Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Struct. Biol.* 23:22–31. https://doi.org/10.1016/j.coisb.2020.08.002.
3. Slattery, M., T. Zhou, …, R. Rohs. 2014. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39:381–399. https://doi.org/10.1016/j.tibs.2014.07.002.
4. Gupta, G., M. Bansal, and V. Sasisekharan. 1980. Conformational flexibility of DNA: polymorphism and handedness. *Proc. Natl. Acad. Sci. USA.* 77:6486–6490. https://doi.org/10.1073/pnas.77.11.6486.
5. Rohs, R., X. Jin, …, R. S. Mann. 2010. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79:233–269. https://doi.org/10.1146/annurev-biochem-060408-091030.
6. Gordân, R., N. Shen, …, M. L. Bulyk. 2013. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* 3:1093–1104. https://doi.org/10.1016/j.celrep.2013.03.014.
7. Yella, V. R., D. Bhimsaria, …, M. Bansal. 2018. Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. *Nucleic Acids Res.* 46:11883–11897. https://doi.org/10.1093/nar/gky1057.
8. Ghoshdastidar, D., and M. Bansal. 2022. Flexibility of flanking DNA is a key determinant of transcription factor affinity for the core motif. *Biophysical J.* https://doi.org/10.1016/j.bpj.2022.08.015.
9. Kribelbauer, J. F., R. E. Loker, …, R. S. Mann. 2020. Context-dependent gene regulation by homeodomain transcription factor complexes revealed by shape-readout deficient proteins. *Mol. Cell.* 78:152–167.e111. https://doi.org/10.1016/j.molcel.2020.01.027.
10. Joshi, R., J. M. Passner, …, R. S. Mann. 2007. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell.* 131:530–543. https://doi.org/10.1016/j.cell.2007.09.024.
11. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1996. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197. https://doi.org/10.1021/ja955032e.
12. Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91. https://doi.org/10.1080/07391102.1988.10506483.
13. Lara-Gonzalez, S., A. C. Dantas Machado, …, C. L. Lawson. 2020. The RNA polymerase α subunit recognizes the DNA shape of the upstream promoter element. *Biochemistry.* 59:4523–4532. https://doi.org/10.1021/acs.biochem.0c00571.
14. Li, J., J. M. Sagendorf, …, R. Rohs. 2017. Expanding the repertoire of DNA shape features for genome-scale studies of transcription factor binding. *Nucleic Acids Res.* 45:12877–12887. https://doi.org/10.1093/nar/gkx1145.
15. Abe, N., I. Dror, …, R. S. Mann. 2015. Deconvolving the recognition of DNA shape from sequence. *Cell.* 161:307–318. https://doi.org/10.1016/j.cell.2015.02.008.
16. Kanhere, A., and M. Bansal. 2005. Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res.* 33:3165–3175. https://doi.org/10.1093/nar/gki627.