Nucleic Acids Research, 2024, **52**, W7–W12
https://doi.org/10.1093/nar/gkae433
Advance access publication date: 27 May 2024
Web Server issue

Deep DNAshape webserver: prediction and real-time visualization of DNA shape considering extended *k*-mers

Jinsen Li [1] **and Remo Rohs** [1,2,3,4,*]

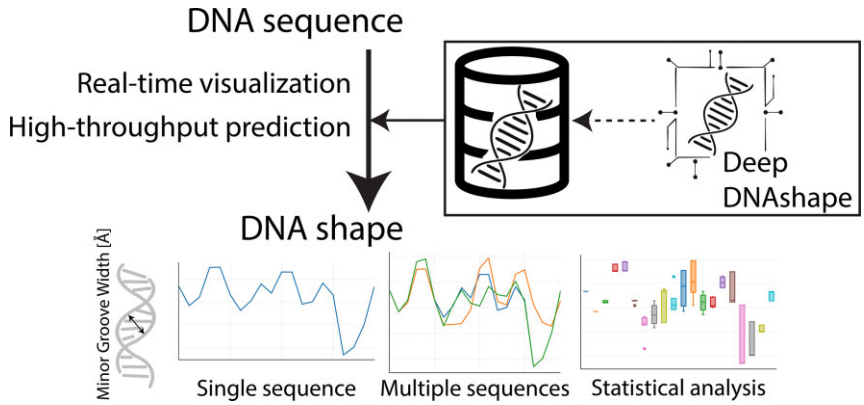
[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA
[2]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA
[3]Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA
[4]Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA

[*]To whom correspondence should be addressed. Tel: +1 213 740 0552; Fax: +1 213 821 4257; Email: rohs@usc.edu

Abstract

Sequence-dependent DNA shape plays an important role in understanding protein–DNA binding mechanisms. High-throughput prediction of DNA shape features has become a valuable tool in the field of protein–DNA recognition, transcription factor–DNA binding specificity, and gene regulation. However, our widely used webserver, DNAshape, relies on statistically summarized pentamer query tables to query DNA shape features. These query tables do not consider flanking regions longer than two base pairs, and acquiring a query table for hexamers or higher-order *k*-mers is currently still unrealistic due to limitations in achieving sufficient statistical coverage in molecular simulations or structural biology experiments. A recent deep-learning method, Deep DNAshape, can predict DNA shape features at the core of a DNA fragment considering flanking regions of up to seven base pairs, trained on limited simulation data. However, Deep DNAshape is rather complicated to install, and it must run locally compared to the pentamer-based DNAshape webserver, creating a barrier for users. Here, we present the Deep DNAshape webserver, which has the benefits of both methods while being accurate, fast, and accessible to all users. Additional improvements of the webserver include the detection of user input in real time, the ability of interactive visualization tools and different modes of analyses. URL: https://deepdnashape.usc.edu

Graphical abstract



Introduction

Protein–DNA interaction is fundamental in many biological processes such as gene regulation. Knowing the structural properties of DNA double-helical molecules is a key step in understanding protein–DNA readout mechanism. Knowing the DNA base sequence is crucial. However, proteins not only interact with DNA base chemical properties (1) but also recognize DNA structure through DNA shape readout (2). Therefore, DNA shape is a critical property that cannot be overlooked (3). Our definition of DNA shape includes the important properties of conformational flexibility (4) and minor-groove electrostatic potential (5). Nevertheless, we are critically aware that local DNA shape is only one scale where genes are regulated (6–8). Additional molecular and cellular scales influence gene regulation in vivo, which involves complex cellular processes such as genome folding (9), nucleosome positioning (10), DNA methylation (11), and histone modification (12). Structural properties of DNA molecules are an important level of studies for revealing molecular mechanisms of protein–DNA binding (13,14) and gene regulation (15,16).

Although DNA shape is sequence-specific, meaning different DNA sequences should have different structural configurations, there are different DNA sequences with similar three-dimensional shape, and likewise single nucleotide polymor-

Received: February 26, 2024. **Revised:** April 30, 2024. **Editorial Decision:** May 5, 2024. **Accepted:** May 8, 2024

phisms can vary DNA structure in an extended region. To reveal structural mechanisms of protein–DNA readout on a large scale, we need quick access to such structural properties. Molecular simulations or experimental methods are time-consuming and not capable for large-scale applications (17). Therefore, the DNAshape webserver came into play (18). In the underlying method, a pentamer query table was created with four DNA shape features. This method was later expanded to a set of 14 DNA shape features including minor-groove electrostatic potential (5,17). The pentamer query table was a summarized statistical average derived from a large quantity of Monte-Carlo simulations of 2,121 DNA fragments of 12–27 base pairs (bp) in length. The DNAshape webserver was well-received in the community and assisted many protein–DNA binding studies (19–28), particularly, in predicting and understanding protein–DNA binding interactions. However, the underlying query tables of the DNAshape webserver have never been expanded to longer *k*-mers due to the exponential computational cost. The DNAshape webserver also has a design with only static input and output and provides basically no real-time interactions with users. An update of the DNAshape webserver and improved usability was urgently needed.

Recent advancement in deep learning enabled the development of a new framework, Deep DNAshape (29), that can predict DNA structure in form of accurate DNA shape features considering extended flanking regions. Even when compared to the DNAshape pentamer query table, the Deep DNAshape method corrected biases and artifacts that may exist in the raw data. However, it requires the downloading of a Python package and installing of multiple frameworks that may not be a good fit for every researcher. For users who only want to inspect the shape profiles of a limited number of sequences, these prerequisites are obstacles for usage. In addition, running a deep learning model requires loading parameters and using advanced computational resources, adding further barriers to the prediction process. A fast webserver with modern amenities will likely attract a larger user base.

Therefore, we present the Deep DNAshape webserver, allowing users to visualize DNA shape features for any nucleotide sequence(s) in real-time. Rather than running a deep learning model in the server backend, the Deep DNAshape webserver utilizes a sliding window algorithm on query tables that are newly generated by the Deep DNAshape method compared to the statistical average over simulation data in the pentamer-based DNAshape webserver. This retrofit expedites the prediction speed. The Deep DNAshape webserver interface was designed with an event-driven real-time prediction of DNA shape features, enabling immediate visualization. In addition, the plotted figures can be customized online in real-time, generating publication-ready figures with just drags and clicks. For users that need to process a file input, the Deep DNAshape webserver provides upload and download functions for offline analyses.

Materials and methods

Deep DNAshape method and predicted DNA shape features

Deep DNAshape method overview

The Deep DNAshape method is a deep learning model that learns structural information directly from structural data sources such as Monte Carlo (MC) simulations, molecular dynamics simulations and experimental sources (29). The method relies on a meticulously designed architecture to infer DNA shape features influenced by different combinations of flanking bp, in a layer-to-layer manner (29). The method can predict DNA shape features by considering up to 7-bp flanking regions for any input DNA sequence containing the letters A, C, G, T or N. The method provides different outputs from different layers considering flanking regions of k bp (with k = 0 to 7). For instance, when $k = 2$, the influence of two bp 5′ and two bp 3′ are considered for the DNA shape at the central bp, resulting in a pentamer-like prediction.

DNA shape features

DNA shape features predicted by Deep DNAshape include six intra-bp features, six inter-bp features, and two minor groove features. Intra-bp features describe translations (Shear, Stretch, and Stagger), and rotations (Buckle, Propeller Twist (ProT) and Opening) between two bases within a single bp. Inter-bp features describe the translations (Shift, Slide, and Rise), and rotations (Tilt, Roll, and Helix Twist (HelT)) between two adjacent bp. The two minor groove features are minor groove width (MGW) and electrostatic potential (EP). Detailed definitions of these parameters can be found in (5,17,29–31).

DNA shape fluctuation features

DNA shape fluctuation features are defined based on the fluctuation (standard deviation) observed in MC simulations of the corresponding DNA shape features. Higher fluctuation values are indicative of flexible regions. In our visualizations provided by the Deep DNAshape webserver, DNA shape fluctuations are shown as error bars on line plots. Users can also download these values for direct value comparisons.

Retrofit query table generation

Inspired by the Deep DNAshape architecture (29) and the pentamer-based DNAshape webserver (18), we generated the complete *k*-mer table ($k = 5$ to 11) containing all possible *k*-mer combinations from A, C, G, T and N and their corresponding DNA shape features at the central bp or bp step. For intra-bp features and minor groove features, odd *k*-mer tables ($k = 5, 7, 9, 11$) were generated. For inter-bp features, even *k*-mer tables ($k = 6, 8, 10$) were generated. For *k* larger than 11, total counts of entries would require an unrealistic amount of storage space. Users are encouraged to use the Deep DNAshape method (29) directly if they aim to study DNA shape features likely influenced by farther flanking regions. For the Deep DNAshape webserver, generated query tables are stored in Apache Parquet format for easy storage and loading.

Server-side light-weight daemon program for database query

The daemon program controlling the databases of the query table (query-program) is pre-loaded when the server is started. Communications between the daemon program and the server backend program happen via a Python remote objects library (Pyro4). The daemon program exposes its functions to the backend program, so the backend program treats these functions as they are native. Different *k*-mer tables were pre-loaded for different layer settings. A sliding window technique is then used to query the desired DNA shape feature from the database.

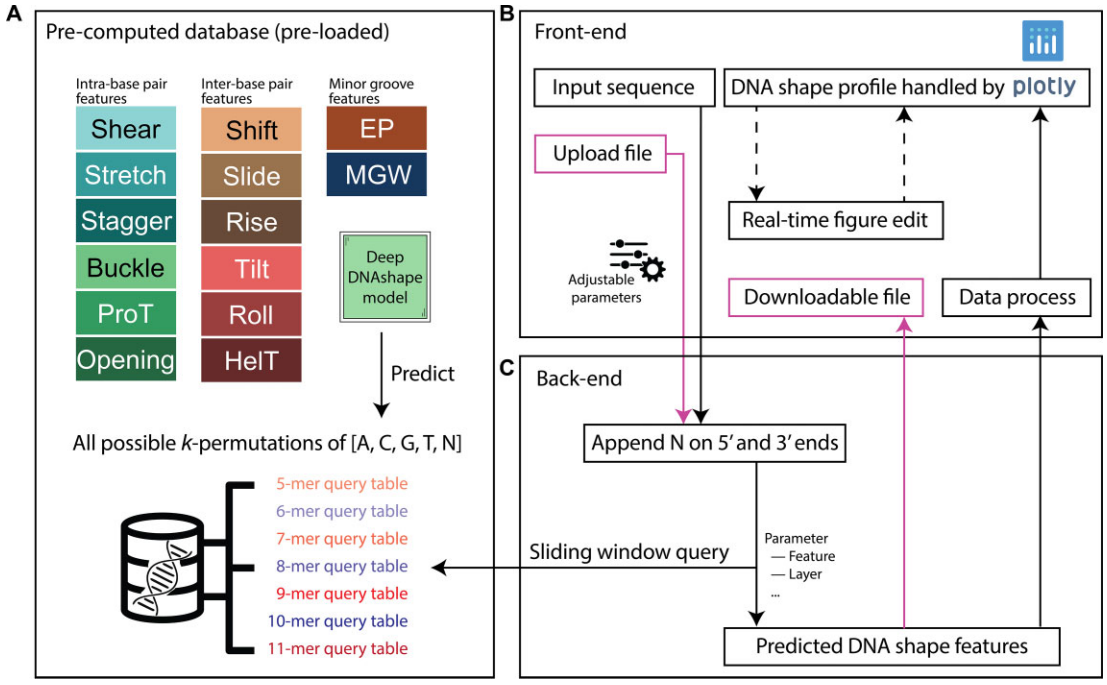


Figure 1. Framework of Deep DNAshape webserver. (**A**) Generation of the database for query tables. The daemon program loads all pre-computed query tables as the webserver starts and exposes its functions of predicting DNA shape features to the backend program. (**B**) Frontend design scheme and functionalities. (**C**) Backend design scheme and functionalities.

Webserver backend and frontend design

The webserver uses a Python FLASK library for its backend. The overall design is demonstrated in Figure 1. The backend handles user requests to illustrate HTML pages from templates in real time. User input of DNA sequences and DNA sequence files are handled in the backend of the webserver. The backend will call functions in the pre-loaded daemon program to retrieve DNA shape features for each DNA sequence. N-caps are added at the 5′ and 3′ ends according to the layer setting to accommodate terminal regions as noted in the above section. Retrieved DNA shape features are returned to the frontend for visualization purposes. The frontend utilizes the Plotly.js package to illustrate the data as line plot or boxplot. Customized buttons for downloading data and downloading the plot as images are added through Plotly.js integration. For users who want to upload separate files and analyze offline, a button to upload a file and download the prediction results is added. The webserver is designed to function straight forward and is easy to use. A manual page is provided to the user through the webpage.

Sliding window algorithm to query DNA shape

A sliding window algorithm (Figure 2) is used for querying all DNA sequences. For a DNA sequence of the form $[X_1, X_2, \ldots, X_n]$ and a k-mer query table, where X_i is either A, C, G, T or N, we will first add N-caps to the sequence to assume the form $[N, \ldots, N, X_1, X_2, \ldots, X_n, N, \ldots, N]$, where the number of N is the same at the 5′ and 3′ ends and either side is $l = \lfloor \frac{k-1}{2} \rfloor$. To query the DNA shape feature of X_i, simply query the database for the subset of the string as $[X_{i-l} \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_{i+l}]$. The returned value will be assigned S_i, the predicted DNA shape feature for X_i. Following iteration through all positions, the returned array will assume the form $[S_1, S_2, \ldots, S_n]$.

Security protocols to ensure data privacy

The webserver uses several safety protocols to protect the users' privacy, ensuring the users' confidence to process sensitive genomic data on the server. The website is secured by an SSL/TLS certificate (HTTPS) distributed by the University of Southern California. The website does not associate IP addresses with users and does not use cookies to identify users' content. When the user types in the textbox for real-time DNA shape profle prediction, neither input or results are tracked nor stored on the server. If the user wants to revisit the DNA shape profile once the webpage is closed, the user needs to reopen the webserver and re-input their sequences. When the user uploads sequence text files or FASTA files, files will be randomly renamed (using a 128-bit universally unique identifier (UUID)) for temporary storage in a dedicated 'upload' folder on the server. Once the prediction of DNA shape profiles is completed and the results are generated in a dedicated 'download' folder on the server, a URL to download the results will be provided and the uploaded file will be deleted. Immediately after the user receives the file, the generated results file will be deleted on the server to ensure maximum security.

Results

Real-time illustration of DNA shape features for any number of DNA sequences

For short DNA sequences, predictions can be made in a matter of milliseconds, allowing real-time visualization of DNA shape features for the first time. The pentamer-based DNAshape webserver (18) required users to type in sequences, which is followed by the return of the plotted figures for each DNA sequence with no adjustment allowed. As technology advanced in recent years, real-time visualization and edits can be implemented in the frontend of the webserver. Thus,

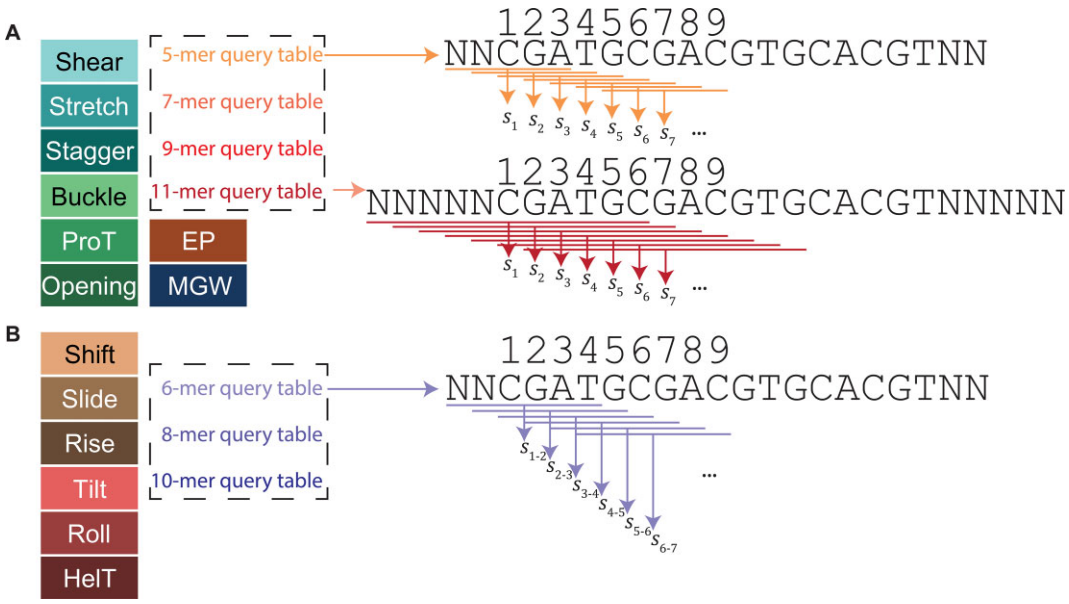


Figure 2. Sliding window algorithm for predicting DNA shape features. (**A**) For intra-base pair (bp) features and minor groove features, odd *k*-mer tables are used because values are assigned to the bp position. (**B**) For inter-bp features, even *k*-mer tables are used because values are assigned to the nucleotide position between consecutive bp.

the Deep DNAshape webserver allows interactive visualization. We completely removed the 'submit' button as the Deep DNAshape webserver detects user input events in the textbox. Typing or editing the input textbox triggers a new DNA shape query in real time. Note that only A, C, G, T and N characters (or their lower-case equivalents) are allowed in the textbox. Other input will prevent the frontend from sending requests to the backend to save computing resources. Real-time updating enables the user to quickly inspect DNA shape profiles of different sequences.

Similar to the pentamer-based DNAshape webserver (18), the Deep DNAshape webserver allows multiple sequences as input, revealing differences in DNA shape profiles between different sequences. Users can also in silico mutate their sequences by editing the sequences in the text box and compare the predicted DNA shape profiles to identify desired mutation targets with similar or different DNA shape, depending on the user's needs. Here, we demonstrate some examples on potential biological applications of the Deep DNAshape webserver (Figure 3).

DNA shape at core region affected by different flanking regions

As shown in Figure 3A, by using the Deep DNAshape web server, MGW profiles of three DNA sequences can be viewed. The position of interest is marked in red, affected by the nearby different flanking regions. The predictions indicate that the difference in DNA shape profile at the core is heavily affected by sequence variations in the nearby flanking regions.

Similar DNA shape of different DNA sequences

Various DNA shape changes can occur for different sequence mutations. As shown in Figure 3C, by using the Deep DNAshape webserver, MGW profiles of two DNA sequences can be visualized. Despite five total mutations that distinguish one sequence from the other, MGW profiles of the two sequences are very similar to each other, indicating a structural similarity between two distinctly different sequences.

Different DNA shape of similar DNA sequences

Compared to Figure 3C, instead of five mutations, only two critical mutations, A–>T and T–>A, will dramatically change the overall MGW profiles between two sequences, as shown in Figure 3D. Users can explore how mutations can affect DNA shape features of their query sequences by simply editing the typed-in nucleotide sequences.

Statistical analysis of bulk sequences

When it comes to protein–DNA readout, it is important to study DNA shape profiles of the core binding site. The frontend plotting framework allows us to change plot type from line plot to boxplot with a click of button. The boxplot shows statistical analysis for a group of sequences in a much cleaner manner compared to line plots, where the distribution of DNA shape preference can be viewed at different DNA positions inside or outside the motif region. We demonstrated an example of boxplot (Figure 3B) for the top 1,000 binding sites of the Max homodimer acquired from a genomic-context protein binding microarray (gcPBM) experiment (32). Although this basic helix-loop-helix (bHLH) protein specifically requires the enhancer box (E-box) 'CACGTG' at the core to be a high affinity binding site, the shape profile at the core can still vary, indicating possible shape readout at the core affected by the flanking regions outside of the E-box.

Offline analysis of DNA shape

For advanced users, text files containing pure DNA sequences or FASTA files can be uploaded. The Deep DNAshape webserver will return a downloadable link containing the predicted DNA shape values based on selected parameters. Users can then incorporate these DNA shape features in their pipeline to tackle different biological problems. For example, users can use the DNA shape features as engineered fea-

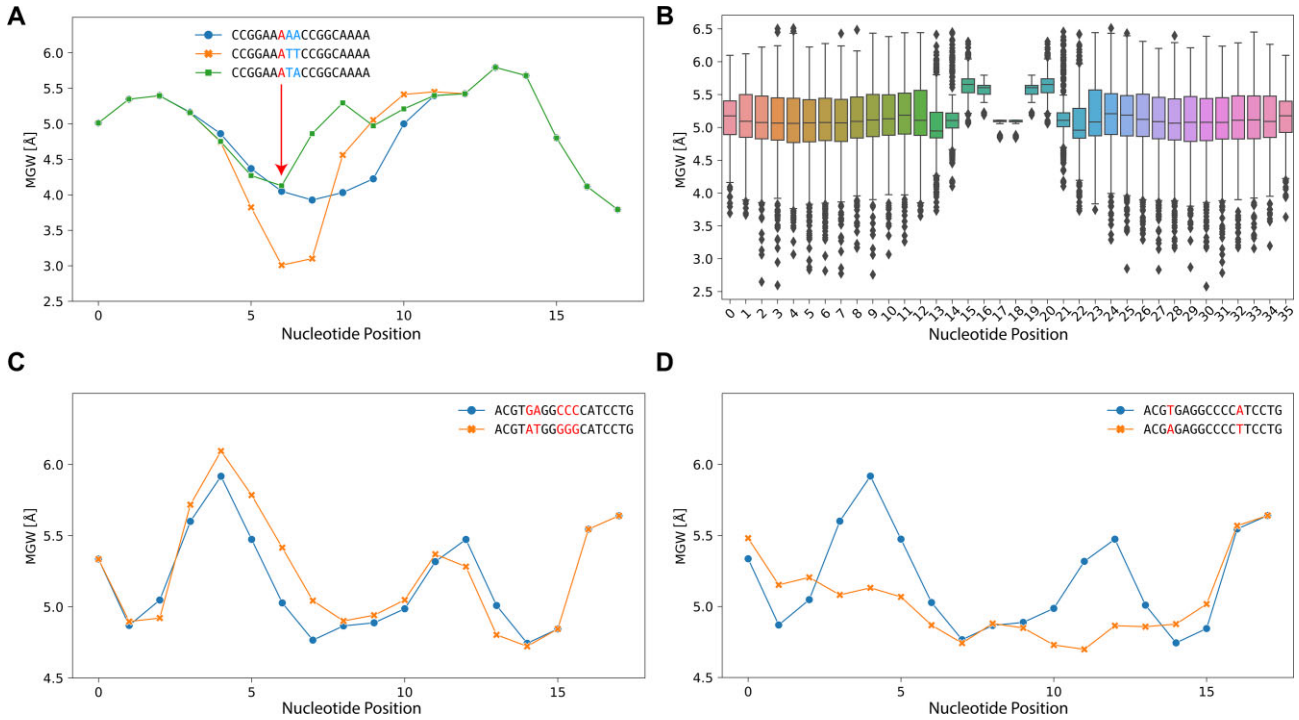


Figure 3. Selected applications of Deep DNAshape webserver. (**A**) Showcase of DNA shape profile at a core affected by different flanking regions. (**B**) Boxplot of top 1,000 binding sites derived from gcPBM experimental data for the Max homodimer. (**C**) Showcase of similar DNA shape profiles of two different DNA sequences. (**D**) Showcase of distinct DNA shape profiles of two similar DNA sequences.

tures for machine learning applications. Users can also directly study the predicted DNA shape features and correlate with the biological problems.

Discussion

Whereas the Deep DNAshape webserver provides a fast and accessible way to immediately predict DNA shape features for any DNA sequence, this method also has limitations. Due to memory and storage restrictions, the webserver can only host a portion of the Deep DNAshape model (29) by generating 5- to 11-mer query tables. As an example, the webserver is designed to predict DNA shape considering 2- to 5-bp flanking regions, compared to the deep learning model with a range from 0- to 7-bp flanking regions 5′ and 3′ of a central bp or bp step. To unlock the full potential of the Deep DNAshape model with the ability to query the influence of longer *k*-mers, users are still required to install the Python package and run the code locally. Furthermore, DNA shape features only describe local DNA structural properties under ideal conditions derived from molecular simulation data. They may not correlate well with global DNA shape in vivo affected by cellular processes that can dramatically alter DNA topology (21,33).

Despite these limitations, the Deep DNAshape webserver will bring value to the user community. Most of all, it will replace the DNAshape webserver (18) by a modern computing approach. From a user perspective, users of the Deep DNAshape webserver have the opportunity to either type in DNA sequences or upload a DNA sequence file, depending on whether they want to analyze their DNA shape data online or offline. For typed-in sequences, production-ready figures will be generated in real-time that are editable directly in the users' browser. For uploaded files, files containing the predicted DNA shape features will be sent to the user for download. The Deep DNAshape webserver removes all current prerequisites and barriers to use the Deep DNAshape model (29) and will unlock much potential to many biological applications for a broad audience.

Data availability

Source code for the Deep DNAshape webserver can be found at https://doi.org/10.6084/m9.figshare.25768959. Predicted *k*-mer query tables can be downloaded at https://doi.org/10.6084/m9.figshare.25286197. Source code of the Deep DNAshape model for generating the *k*-mer query tables can be found at https://doi.org/10.5281/zenodo.10403299 (29).

Acknowledgements

The authors acknowledge comments and suggestions from Rohs Lab members in the process of designing the Deep DNAshape webserver, Tianyin Zhou for the development and validation of the pentamer-based DNAshape webserver (18), and Luigi Mann for setup and maintenance of the Deep DNAshape webserver.

Funding

National Institutes of Health [R35GM130376 to R.R.] and Human Frontier Science Program [RGP0021/2018 to R.R.]. Funding for open access charge: National Institutes of Health [R35GM130376].

Conflicts of interest statement

None declared.

References

1. Chiu,T.P., Rao,S. and Rohs,R. (2023) Physicochemical models of protein–DNA binding with standard and modified base pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2205796120.
2. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein–DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
3. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein–DNA recognition. *Nature*, **461**, 1248–1253.
4. Jiang,Y., Chiu,T.P., Mitra,R. and Rohs,R. (2024) Probing the role of the protonation state of a minor groove-linker histidine in Exd-Hox–DNA binding. *Biophys. J.*, **123**, 248–259.
5. Chiu,T.P., Rao,S., Mann,R.S., Honig,B. and Rohs,R. (2017) Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein–DNA binding. *Nucleic Acids Res.*, **45**, 12565–12576.
6. Levo,M. and Segal,E. (2014) In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.*, **15**, 453–468.
7. Inukai,S., Kock,K.H. and Bulyk,M.L. (2017) Transcription factor–DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.*, **43**, 110–119.
8. Zeitlinger,J. (2020) Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Syst. Biol.*, **23**, 22–31.
9. Davidson,I.F. and Peters,J.-M. (2021) Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.*, **22**, 445–464.
10. Kornberg,R.D. and Lorch,Y. (2020) Primary role of the nucleosome. *Mol. Cell*, **79**, 371–375.
11. Mattei,A.L., Bailly,N. and Meissner,A. (2022) DNA methylation: a historical perspective. *Trends Genet.*, **38**, 676–707.
12. Zhang,Y., Sun,Z., Jia,J., Du,T., Zhang,N., Tang,Y., Fang,Y. and Fang,D. (2021) Overview of Histone Modification. In: Fang,D. and Han,J. (eds.) *Histone Mutations and Cancer. Advances in Experimental Medicine and Biology*. Vol. **1283**, Springer, Singapore.
13. Lawson,C.L., Berman,H.M., Chen,L., Vallat,B. and Zirbel,C.L. (2023) The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids. *Nucleic Acids Res.*, **52**, D245–D254.
14. Baek,M., McHugh,R., Anishchenko,I., Jiang,H., Baker,D. and DiMaio,F. (2024) Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods*, **21**, 117–121.
15. Abe,N., Dror,I., Yang,L., Slattery,M., Zhou,T., Bussemaker,H.J., Rohs,R. and Mann,R.S. (2015) Deconvolving the recognition of DNA shape from sequence. *Cell*, **161**, 307–318.
16. Poul,Y.L., Xin,Y., Ling,L., Mühling,B., Jaenichen,R., Hörl,D., Bunk,D., Harz,H., Leonhardt,H., Wang,Y., *et al.* (2020) Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci. Adv.*, **6**, eabe2955.
17. Li,J., Sagendorf,J.M., Chiu,T.P., Pasi,M., Pérez,A. and Rohs,R. (2017) Expanding the repertoire of DNA shape features for genome-scale studies of transcription factor binding. *Nucleic Acids Res.*, **45**, 12877–12887.
18. Zhou,T., Yang,L., Lu,Y., Dror,I., Dantas Machado,A.C., Ghane,T., Di Felice,R. and Rohs,R. (2013) DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.*, **41**, W56–W62.
19. Wang,S., Zhang,Q., Shen,Z., He,Y., Chen,Z.-H., Li,J. and Huang,D.-S. (2021) Predicting transcription factor binding sites using DNA shape features based on shared hybrid deep learning architecture. *Mol. Ther. - Nucleic Acids*, **24**, 154–163.
20. Horton,C.A., Alexandari,A.M., Hayes,M.G.B., Marklund,E., Schaepe,J.M., Aditham,A.K., Shah,N., Suzuki,P.H., Shrikumar,A., Afek,A., *et al.* (2023) Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*, **381**, eadd1250.
21. Basu,A., Bobrovnikov,D.G., Cieza,B., Arcon,J.P., Qureshi,Z., Orozco,M. and Ha,T. (2022) Deciphering the mechanical code of the genome and epigenome. *Nat. Struct. Mol. Biol.*, **29**, 1178–1187.
22. Lee,C.S.K., Cheung,M.F., Li,J., Zhao,Y., Lam,W.H., Ho,V., Rohs,R., Zhai,Y., Leung,D. and Tye,B.-K. (2021) Humanizing the yeast origin recognition complex. *Nat. Commun.*, **12**, 33.
23. Sielemann,J., Wulf,D., Schmidt,R. and Brautigam,A. (2021) Local DNA shape is a general principle of transcription factor binding specificity in *Arabidopsis thaliana*. *Nat. Commun.*, **12**, 6549.
24. Li,Y., Kong,F., Cui,H., Wang,F., Li,C. and Ma,J. (2023) SENIES: DNA shape enhanced two-layer deep learning predictor for the identification of enhancers and their strength. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **20**, 637–645.
25. Yang,J., Ma,A., Hoppe,A.D., Wang,C., Li,Y., Zhang,C., Wang,Y., Liu,B. and Ma,Q. (2019) Prediction of regulatory motifs from human Chip-sequencing data using a deep learning framework. *Nucleic Acids Res.*, **47**, 7809–7824.
26. Zeiske,T., Baburajendran,N., Kaczynska,A., Brasch,J., Palmer,A.G., Shapiro,L., Honig,B. and Mann,R.S. (2018) Intrinsic DNA shape accounts for affinity differences between hox-cofactor binding sites. *Cell Rep.*, **24**, 2221–2230.
27. Kribelbauer,J.F., Loker,R.E., Feng,S., Rastogi,C., Abe,N., Rube,H.T., Bussemaker,H.J. and Mann,R.S. (2020) Context-dependent gene regulation by homeodomain transcription factor complexes revealed by shape-readout deficient proteins. *Mol. Cell*, **78**, 152–167.
28. Pal,S., Hoinka,J. and Przytycka,T.M. (2019) Co-SELECT reveals sequence non-specific contribution of DNA shape to transcription factor binding in vitro. *Nucleic Acids Res.*, **47**, 6632–6641.
29. Li,J., Chiu,T.P. and Rohs,R. (2024) Predicting DNA structure using a deep learning method. *Nat. Commun.*, **15**, 1243.
30. Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.
31. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
32. Gordân,R., Shen,N., Dror,I., Zhou,T., Horton,J., Rohs,R. and Bulyk,M.L. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.*, **3**, 1093–1104.
33. Basu,A., Bobrovnikov,D.G., Qureshi,Z., Kayikcioglu,T., Ngo,T.T.M., Ranjan,A., Eustermann,S., Cieza,B., Morgan,M.T., Hejna,M., *et al.* (2021) Measuring DNA mechanics on the genome scale. *Nature*, **589**, 462–467.

Received: February 26, 2024. **Revised:** April 30, 2024. **Editorial Decision:** May 5, 2024. **Accepted:** May 8, 2024