SUPPLEMENTARY DATA

Landscape of DNA binding signatures of myocyte enhancer factor-2B reveals a unique interplay of base and shape readouts

Ana Carolina Dantas Machado[1], Brendon Cooper[1], Xiao Lei[2], Rosa Di Felice[1,3], Lin Chen[2,4,5], and Remo Rohs[1,3,4,5,6,*]

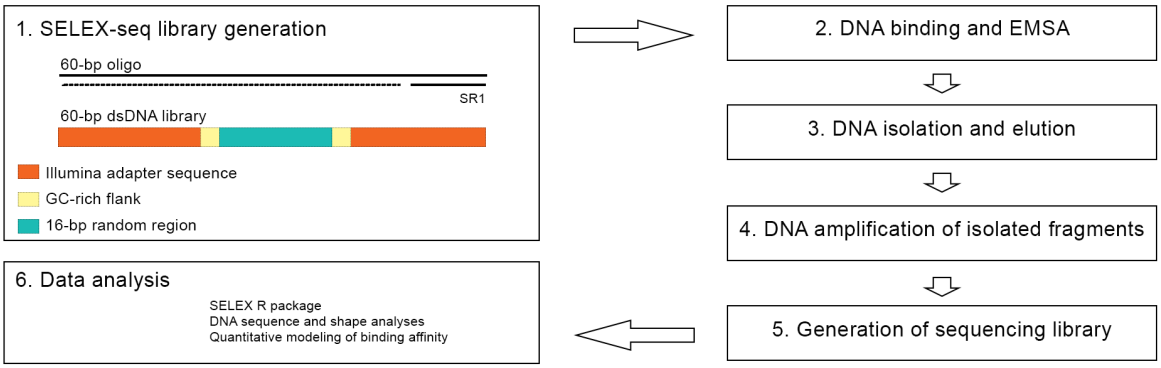
[1]Quantitative and Computational Biology, [2]Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA
[3]Department of Physics & Astronomy, University of Southern California, Los Angeles, CA 90089, USA
[4]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA
[5]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA
[6]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

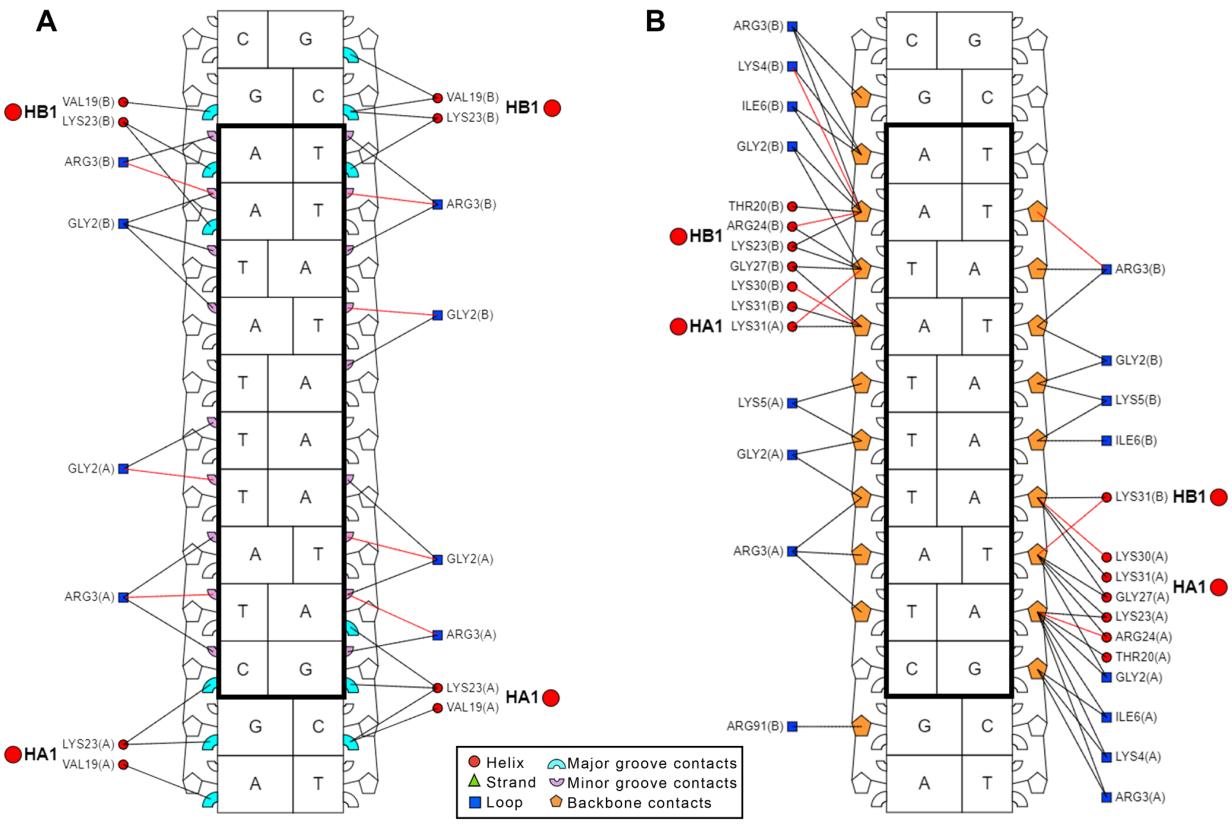
Present address: Ana Carolina Dantas Machado, Department of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Present address: Xiao Lei, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

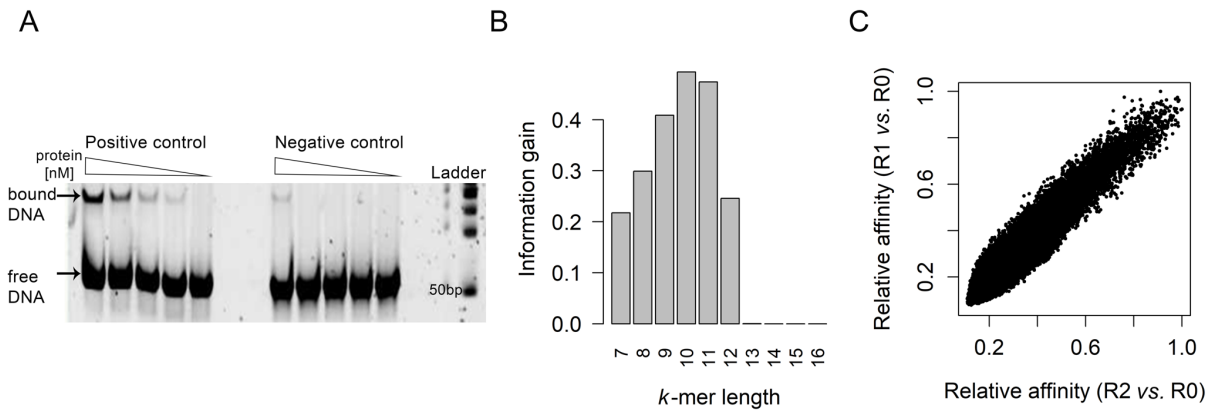
SUPPLEMENTARY FIGURES



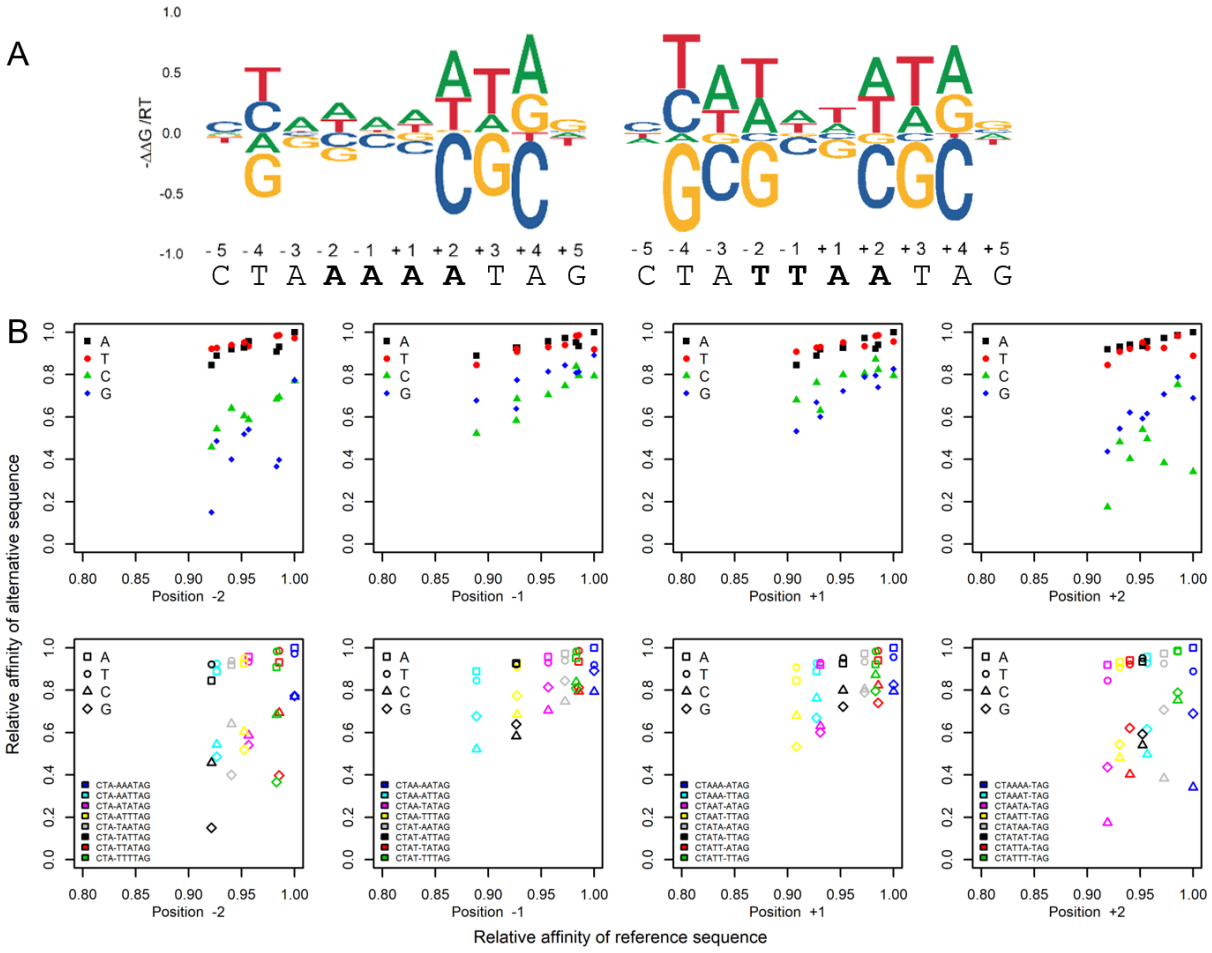
Supplementary Figure S1. Overview of SELEX-seq approach. DNA library used for SELEX-seq experiments contained a 16-bp random region surrounded by flanks and Illumina adapter regions. In each round of selection, the "selected" library was generated by performing Steps 2–4. Selected libraries were used in the next round of selection or prepared for sequencing. All steps were based on a previously described protocol (1, 2). Data were analyzed by using available tools described in the Materials and Methods section.



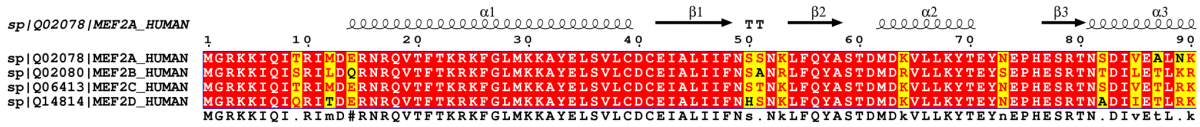
Supplementary Figure S2. Nucleotide-residue interaction maps of MEF2B–DNA complex. **A)** Contacts with major and minor grooves of the DNA. **B)** Contacts with DNA backbone. Regions encircled by bold black box represent the 10-bp core binding site, corresponding to nucleotide positions –5 to +5 based on consensus motif shown in Figure 1. Nucleotide-residue map was generated with DNAproDB (3, 4), based on analysis of the MEF2B–DNA co-crystal structure with PDB ID 1N6J (5).



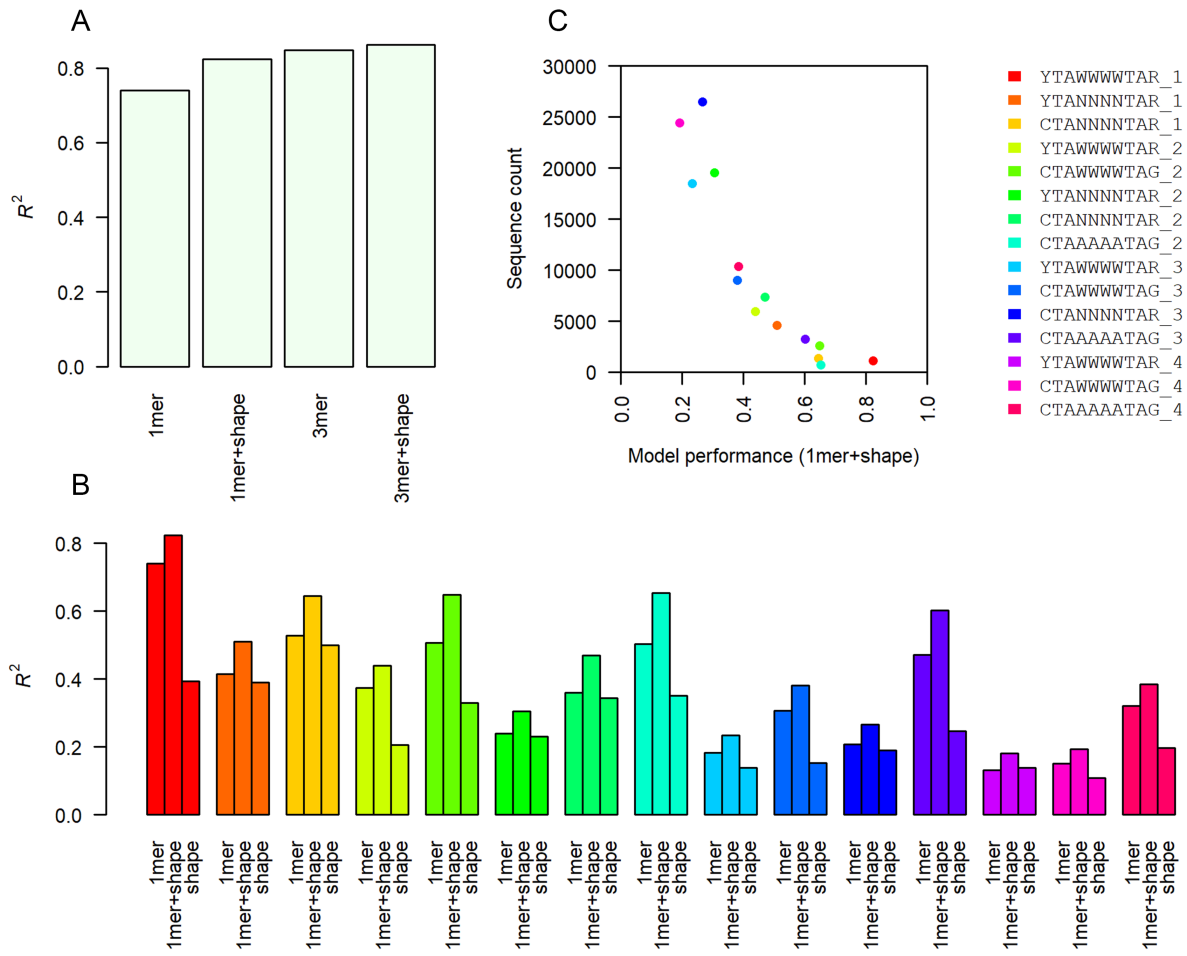
Supplementary Figure S3. SELEX-seq data for MEF2B. A) Validation of library design using positive and negative controls for EMSA. At lower concentrations, using the conditions described in Materials and Methods, MEF2B binds to the positive but not the negative control. Final MEF2B protein concentration was 0, 11, 22, 44, or 88 nM for a fixed amount of control DNA probe (35 nM). **B)** Information gain after two rounds of selection was greatest for 10-mers. Therefore, we based our analysis on sequences of this length. **C)** Relative binding affinity comparison of 10-mers between rounds 1 and 2 of selection.



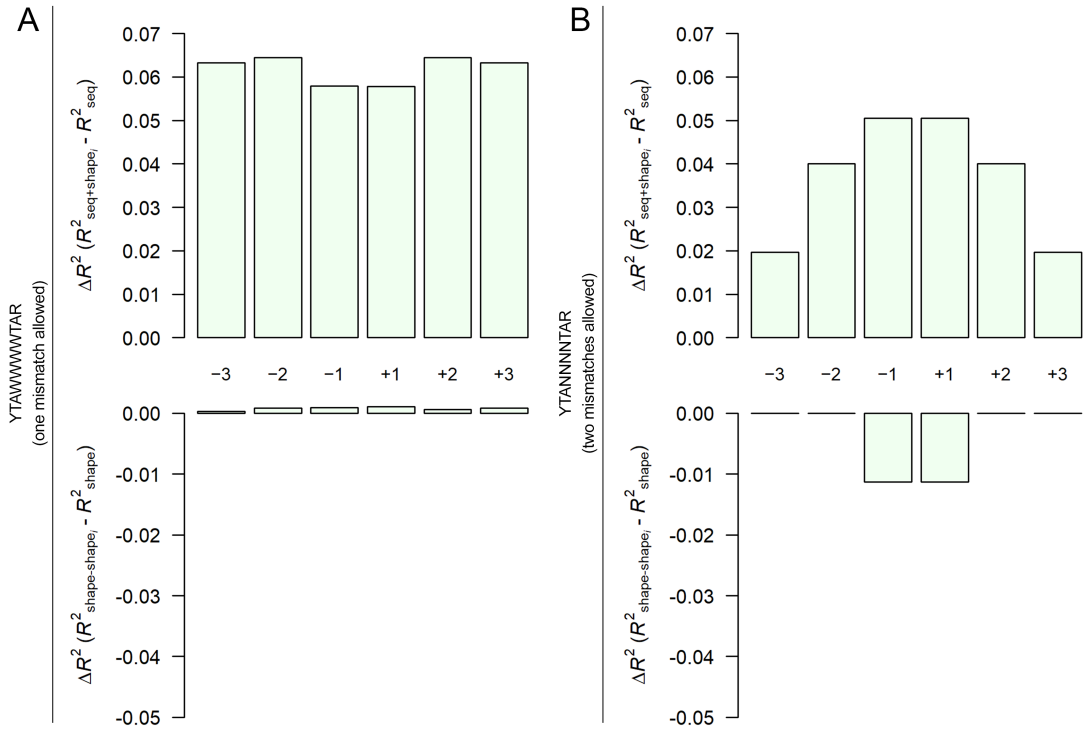
Supplementary Figure S4. Distinct effects of nucleotide substitutions across MEF2B binding site. A) Affinity logos generated based on reference sequence show effects of single-nucleotide substitutions on MEF2B binding sites (logo in left panel appears in Figure 2A and is shown here for comparison). **B)** Comparison of relative affinities for reference (highest affinity) and alternative (lower affinity) sequences based on core motif (CTAWWWWTAR) and variations of the motif at each of the central four nucleotide positions (–2, –1, +1, +2). The same data are displayed on the top and bottom panels at each position, colored differently to emphasize the relative affinity comparisons of distinct features. Top panels highlight nucleotide substitutions for each position within the central core. Bottom panels highlight nucleotide substitutions for each position within the binding site for each specific sequence.



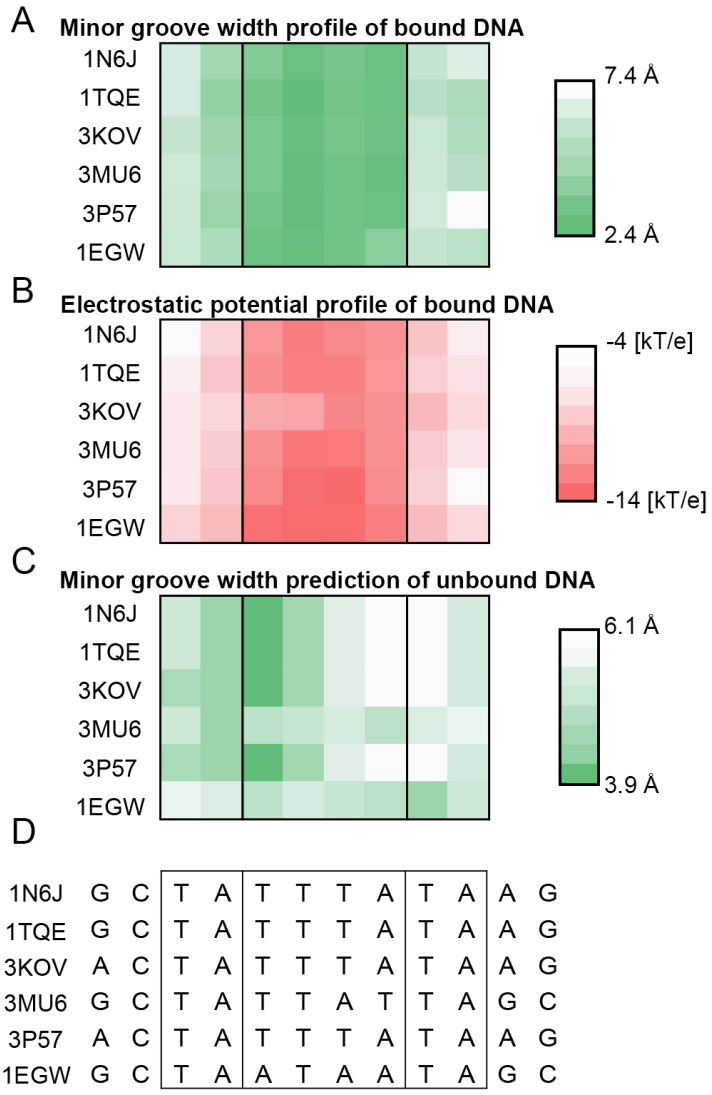
Supplementary Figure S5. Multiple sequence alignment. Protein sequence alignment of the DBD of human MEF2 proteins (residues 1–90): MEF2A (Q02078), MEF2B (Q02080), MEF2C (Q06413), and MEF2D (Q14814). Identifiers are UniProt accession numbers. Sequence alignment was performed with T-Coffee (6) and visualized with ESPript 3.0 (7). Yellow highlights indicate amino acid positions where residues varied among the MEF2 proteins.



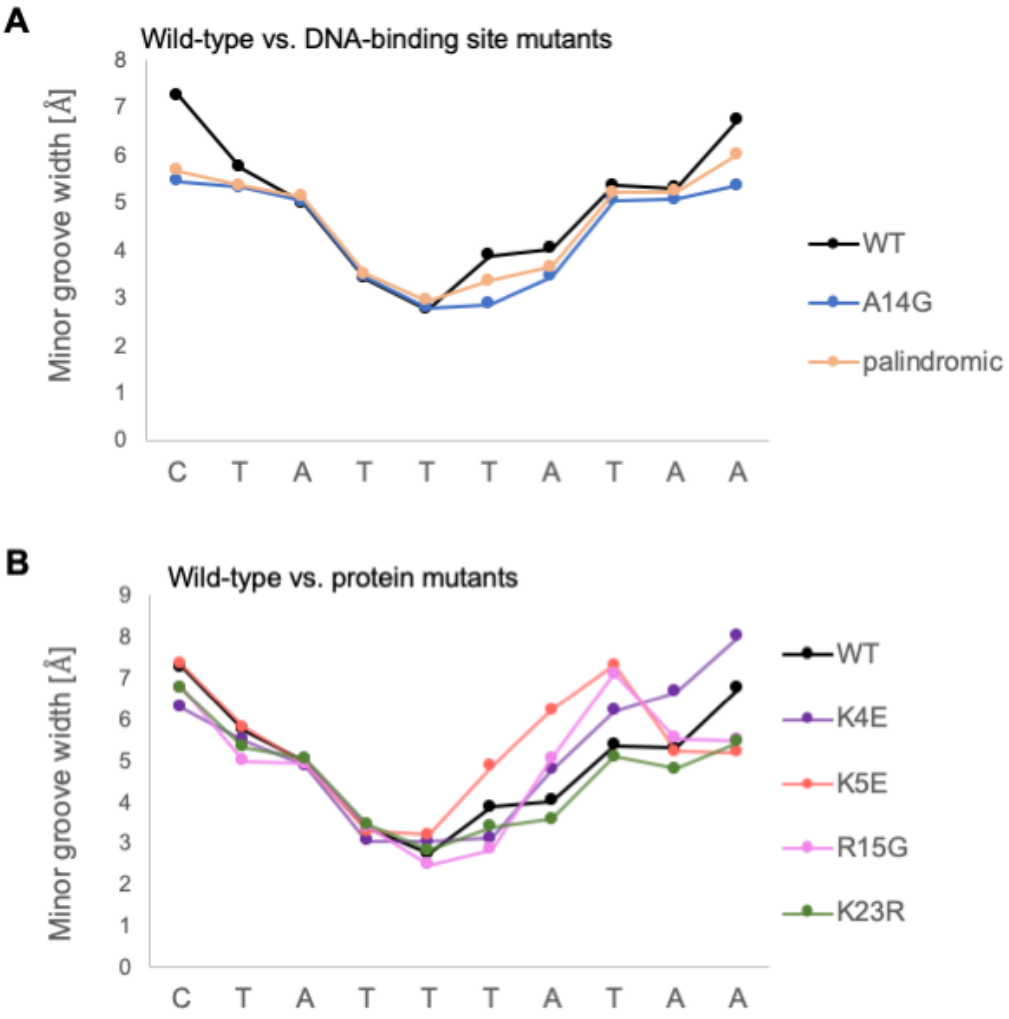
Supplementary Figure S6. Model performance comparison. A) Model performance for DNA sequence and shape features when considering the consensus motif. **B)** Model performance comparison between sequence-only mono-nucleotide (1mer), shape-augmented (1mer+shape), and shape-only (shape) models for multiple datasets with various filtering schemes. Addition of shape features increases model performance regardless of DNA sequence motif used for filtering input sequences in MLR analysis. **C)** Model performance as a function of the number of sequences used. Color scheme is the same as shown in panel B. Sequences used in MLR analysis were selected based on motif and mismatch number shown.



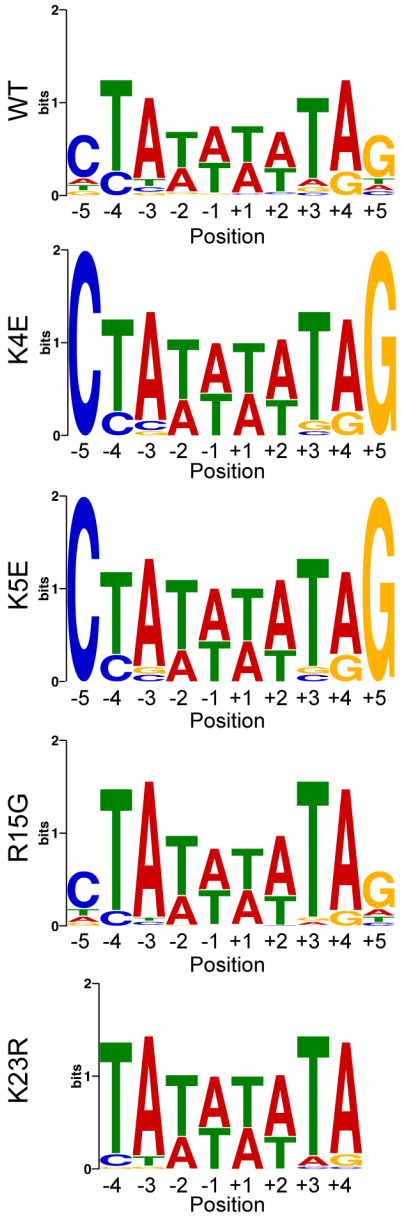
Supplementary Figure S7. Feature selection. Importance of DNA shape for model performance, according to sequence composition when considering **A)** an AT-rich central core or **B)** any composition at the central core. Top panel: Effect on performance (ΔR^2) when shape is added one position at a time to a sequence-only mono-nucleotide model. Bottom panel: Effect on performance (ΔR^2) when shape is removed one position at a time from a shape-only model. Data shown in (B) are the same as shown in the main Figure 4C, and presented here for comparison.



Supplementary Figure S8. Minor groove width and electrostatic potential profiles of MEF2 proteins. A-C) Minor groove width and electrostatic potential profiles of DNA fragments in complex with MEF2, where (A) minor groove width of bound DNA, (B) electrostatic potential of bound DNA, and (C) minor groove width prediction of unbound DNA are shown for **D)** DNA in complexes from PDB IDs 1N6J (5), 1TQE (8), 3KOV (9), 3MU6 (10), 3P57 (11), and 1EGW (12)
.



Supplementary Figure S9. Minor groove width (MGW) profiles derived from MD simulations. **A)** Shape of the bound DNA when mutations on the DNA binding site are introduced ('A14G' refers to a DNA where peripheral half-sites are palindromic: 5'-CTATTTATAG; 'palindromic' refers to a fully palindromic site 5'-CTATTAATAG). **B)** Shape of the bound DNA when protein mutations are introduced.



Supplementary Figure S10. Position weight matrices (PWMs) of MEF2B proteins. PWMs were obtained from SELEX-seq data for *k*-mers with relative affinity > 0.7 and were generated using MEME Suite as described in the methods section. PWMs are shown for wild-type (WT, also shown in main Figure 1 and re-examined here for comparison) and mutant proteins K4E, K5E, R15G and K23R.

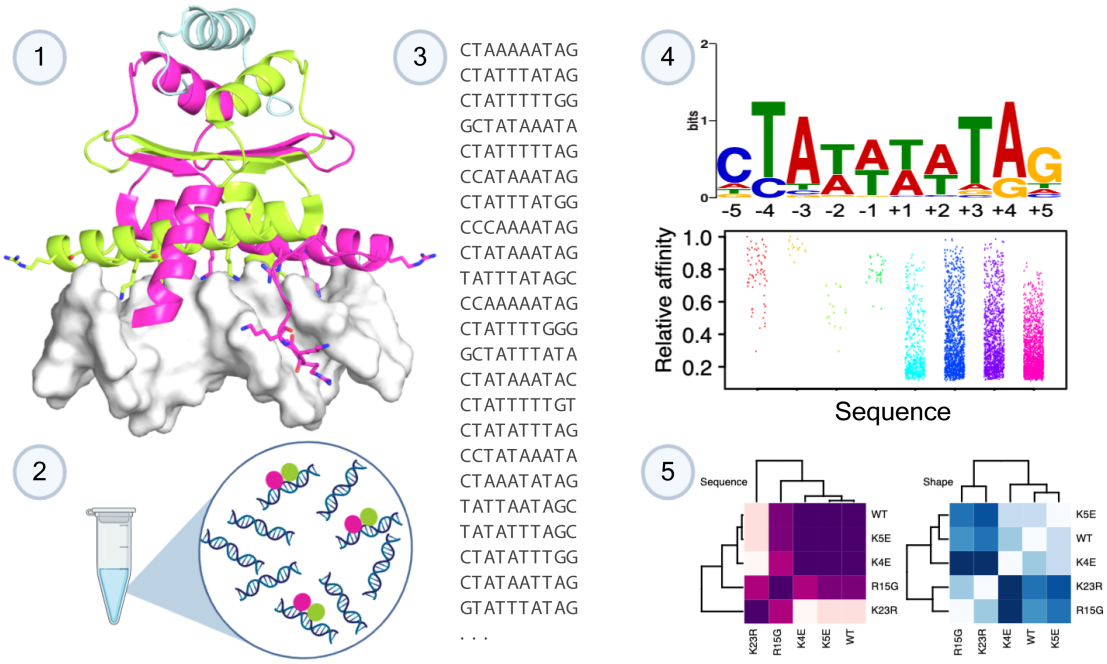
SUPPLEMENTARY TABLE

Supplementary Table S1. Oligonucleotide (oligo) sequences

Oligo ID	Oligo Sequence (5' - 3')
Selex-Lib	GAGTTCTACAGTCCGACGATCCGC[N_{16}]CCTGGAATTCTCGGGTGCCA
SR1	TGGCACCCGAGAATTCCA
SF1	GAGTTCTACAGTCCGACGAT
SR1-FAM	/56-FAM/TGGCACCCGAGAATTCCA
RP1	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA
RPI#	CAAGCAGAAGACGGCATACGAGAT[Index]GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
PCTRL_F	GAGTTCTACAGTCCGACGATCCGCATCTTATAAATAGTCTCCTGGAATTCTCGGGTGCCA
NCTRL_F	GAGTTCTACAGTCCGACGATCCGCATGGTGTGGCTGGTCTCCTGGAATTCTCGGGTGCCA
PCTRL_R	TGGCACCCGAGAATTCCAGGAGACTATTTATAAGATGCGGATCGTCGGACTGTAGAACTC
NCTRL_R	TGGCACCCGAGAATTCCAGGAGACCAGCCACACCATGCGGATCGTCGGACTGTAGAACTC

Abbreviations: PCTRL, positive control; NCTRL, negative control.

GRAPHICAL ABSTRACT



A multi-scale approach to the intricate mode of MEF2–DNA recognition reveals the usage of sequence and shape readout. Graphical abstract was created with BioRender.com. The shown co-crystal structure is PDB ID 1N6J.

AUTHOR CONTRIBUTIONS

A.C.D.M., L.C., and R.R. conceived the study. A.C.D.M. independently led and executed experimental and computational aspects of the project, designed SELEX-seq experiments, and analyzed data. A.C.D.M. and B.H.C. ran and analyzed MD simulations with help from R.D.F. L.X. and L.C. assisted and advised A.C.D.M. in protein purification and binding experiments. A.C.D.M. wrote the paper with B.H.C. and R.R. and help from all authors. R.R. and L.C. supervised the project.

SUPPLEMENTARY REFERENCES

1. Riley,T.R., Slattery,M., Abe,N., Rastogi,C., Liu,D., Mann,R.S. and Bussemaker,H.J. (2014) SELEX-seq: a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods Mol. Biol.*, **1196**, 255–278.

2. Slattery,M., Riley,T., Liu,P., Abe,N., Gomez-Alcala,P., Dror,I., Zhou,T., Rohs,R., Honig,B., Bussemaker,H.J., *et al.* (2011) Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. *Cell*, **147**, 1270–1282.

3. Sagendorf,J.M., Berman,H.M. and Rohs,R. (2017) DNAproDB: an interactive tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.*, **45**, W89–W97.

4. Sagendorf,J.M., Markarian,N., Berman,H.M. and Rohs,R. (2020) DNAproDB: an expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.*, **48**, D277–D287.

5. Han,A., Pan,F., Stroud,J.C., Youn,H.-D., Liu,J.O. and Chen,L. (2003) Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2. *Nature*, **422**, 730–734.

6. Di Tommaso,P., Moretti,S., Xenarios,I., Orobitg,M., Montanyola,A., Chang,J.-M., Taly,J.-F. and Notredame,C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.*, **39**, W13-W17.

7. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320-W324.

8. Han,A., He,J., Wu,Y., Liu,J.O. and Chen,L. (2005) Mechanism of Recruitment of Class II Histone Deacetylases by Myocyte Enhancer Factor-2. *J. Mol. Biol.*, **345**, 91–102.

9. Wu,Y., Dey,R., Han,A., Jayathilaka,N., Philips,M., Ye,J. and Chen,L. (2010) Structure of the MADS-box/MEF2 Domain of MEF2A Bound to DNA and Its Implication for Myocardin Recruitment. *J. Mol. Biol.*, **397**, 520–533.

10. Jayathilaka,N., Han,A., Gaffney,K.J., Dey,R., Jarusiewicz,J.A., Noridomi,K., Philips,M.A., Lei,X., He,J., Ye,J., *et al.* (2012) Inhibition of the function of class IIa HDACs by blocking their interaction with MEF2. *Nucleic Acids Res.*, **40**, 5378–5388.

11. He,J., Ye,J., Cai,Y., Riquelme,C., Liu,J.O., Liu,X., Han,A. and Chen,L. (2011) Structure of P300 Bound to MEF2 on DNA Reveals a Mechanism of Enhanceosome Assembly. *Nucleic Acids Res.*, **39**, 4464–4474.

12. Santelli,E. and Richmond,T.J. (2000) Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution. *J. Mol. Biol.*, **297**, 437–449.