SUPPLEMENTAL INFORMATION

Differential Sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains

May 18, 2016

Dominique C. Stephens[1] and Gregory M. K. Poon[1,2,*]

[1] Department of Chemistry, Georgia State University, Atlanta, GA 30303

[2] Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30303

* To whom correspondence should be addressed at: P.O. Box 3965, Atlanta, GA 30302-3965. Email: gpoon@gsu.edu. Tel. (404) 413-5491.

SUPPLEMENT INDEX **PAGE**

SUPPLEMENTAL METHODS

Generation of duplex SC1 binding sites. Coordinates for a duplex DNA sequence containing the exact sequence as the competitor (5'-CGGCCAAGCCCGGAAGTGAGTGCC-3', harboring the SC1 site in red) used in our titration experiments was constructed in PDB format with fiber B-DNA geometry using 3DNA (1). This unmethylated construct was then derivatized to the hemi- and fully methylated duplexes using the "mutate_base" function in 3DNA.

All-atom energy minimization. The synthetic PDB coordinates were prepared for energy minimization with GROMACS 5.1.2. The coordinates were processed to remove terminal phosphates and parameterized for CHARMM36 (2) using previously published topology for 5-methylcytosine (3). Each structure was centered in a 10 × 10 × 10 nm cubic box, hydrated with TIP3P water at normal density, and neutralized with 0.15 M NaCl. Energy minimization was performed by steepest descent; the maximum force at convergence in all cases was below 1000 kJ mol^{-1} nm^{-1}. Convergence was additionally confirmed by two methods. First, total potential energy showed a gradient of less than 0.01% at termination (**Figure S4**). Second, pairwise alignment of the final structures from four independently prepared simulations gave RMSD values below <0.001.

Analysis of helical parameters. DNA coordinates were extracted from the energy-minimized ensemble and processed with Curves+ (4) to fit helical parameters. 5-methylcytosines were treated exactly as cytosines since the program considered only the ring atoms for least-square fitting, and comparison of the cytosine geometry used by Curves+ was indistinguishable from that for 5-methylcytosine used by 3DNA, which generated the pre-minimized structures. Fitted helical

parameters were categorized as base-centric (intra-bp: shear, stretch, stagger, buckle, propeller, and opening; inter-bp: shift, slide, rise, tilt, roll, twist), backbone (for each strand, phosphate and glycosidic torsional angles α, β, γ, δ, ε, ζ, χ as well as the sugar pseudo-rotation phase and amplitude), and groove widths and depths, or bp-axis. In all, 39 numerical parameters were collected for each structure at each base step, with the exception of groove geometries which were fitted every half base step. In some cases, values at the termini were ill-defined. For groove widths and depths, up to three terminal residues were skipped.

For each parameter, the values along the DNA sequence for the un-, hemi-, and fully methylated site were subject to principal component analysis (PCA) computed with covariance matrices using Origin 9.1. The calculated eigenvalues were used in the main text for **Figure 5A** and Bartlett's test for equality of variances, and the eigenvectors were used to construct the loading plots in **Table 2** in the main text.

Computation of electrostatics. Continuum electrostatic properties of the energy-minimized structures were computed via the Poisson-Boltzmann (PB) formalism using APBS (5). For each structure, coordinates in PDB format were first parameterized for CHARMM (which provides parameterization for 5-methylcytosine) using PDB2PQR (6) and then solved by APBS under the following conditions: non-linear PB equation, 0.15 Na^+ and Cl^- (ionic radii 2.0 and 1.8 Å, respectively), temperature at 298 K, and all other parameters at default. Surface electrostatic potential was plotted using PyMOL. Matched per-atom electrostatic energies for were taken as input for PCA.

INDEX OF SUPPLEMENTAL FIGURES

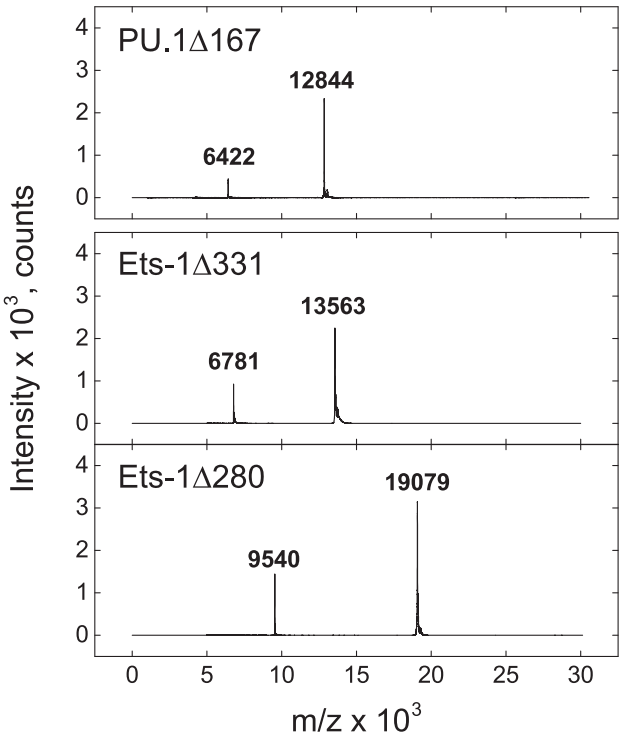

FIGURE S1. **Mass spectrometric analysis of purified recombinant PU.1 and Ets-1 ETS domains.** Eight pmol of protein was mixed with matrix at a 1:1 v/v ratio and analyzed by MALDI-ToF mass spectrometry in linear positive-ion mode. Peaks represent the +1 and +2 ions. PU.1ΔN167: expected 12,978, observed 12,844; Ets-1ΔN331: expected 13,561, observed 13,563; Ets-1ΔN280: expected 19,088, observed 19,079. The expected MW for PU.1 exceeds the observed value by 134, which is consistent with post-translational removal of the N-terminal methionine (-131) *in vivo*. Hydrolytic cleavage by the *E. coli* methionine aminopeptidase is common with Gly in the P1' position (7), which is the case for PU.1ΔN167.

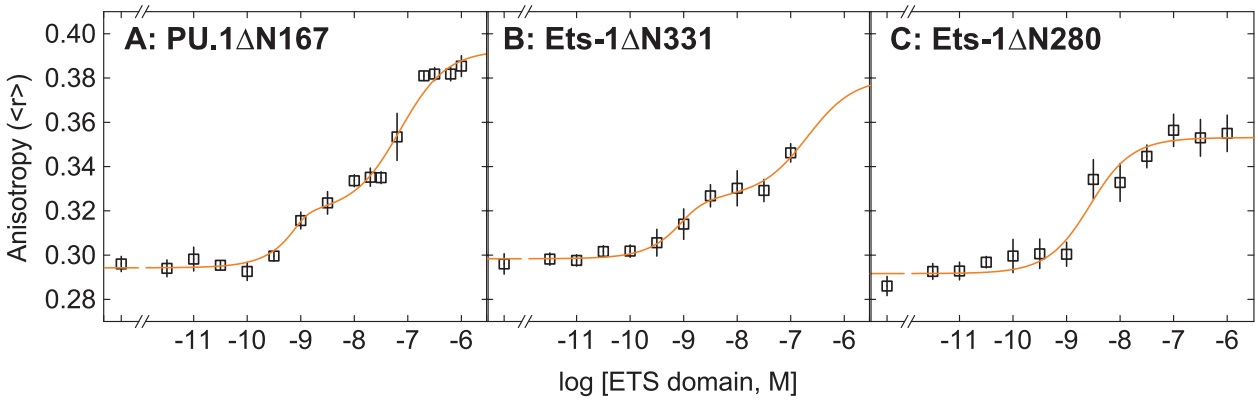

FIGURE S2. **Direct titration of ETS domains with a Cy3-labeled DNA probe.** Binding of ETS domains to un-, hemi-, and fully methylated SC1 was measured by a competitive titration of protein bound to a fluorescent Cy3-labeled DNA probe, whose spectral and PU.1-binding properties were previously described (8). Here, binding to Ets-1 was investigated under identical physiologic conditions as PU.1. In the presence of 1 nM probe, like PU.1Δ167, Ets-1Δ331 bound this probe with high affinity and exhibited a second, negatively cooperative binding of a second equivalent at high concentrations. The probe (1 nM) and protein (2 nM) concentrations chosen for the competitive titrations as shown in **Figure 1** and **Table 1** in the main text were based on the direct binding data expressly to avoid potential complications of ETS:DNA = 2:1 binding. It can be seen that at 2 nM protein ($10^{-8.7}$ M), ETS/DNA binding was quantitatively in the 1:1 régime. The equilibrium dissociation constants for the three constructs to this probe are (mean ± S.E. of *N* replicate experiments in parentheses):

K_D, nM		
PU.1Δ167	Ets-1Δ331	Ets-1Δ331
0.078 ± 0.008 (3)	0.28 ± 0.01 (3)	3.0 ± 1.0 (3)

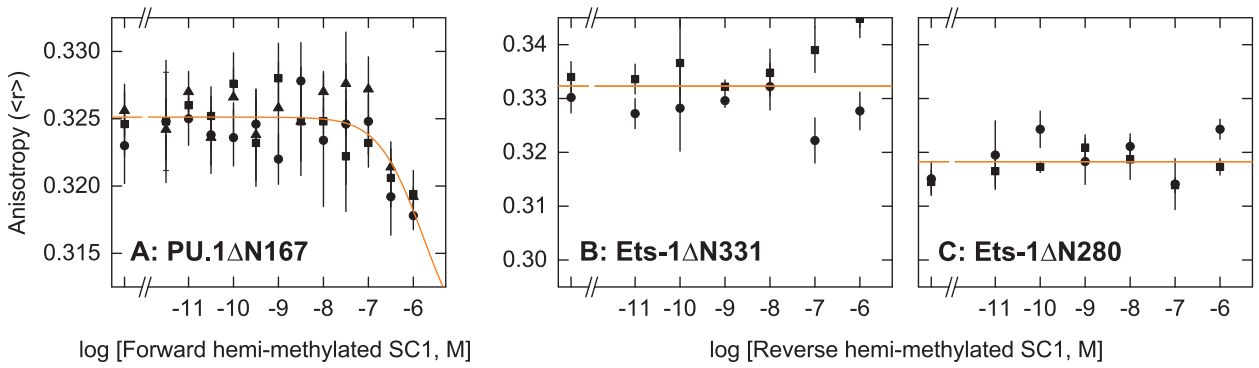

FIGURE S3. **Titrations of the PU.1 and Ets-1 ETS domain with hemi-methylated DNA.** PU.1ΔN167 bound hemi-methylated SC1 site at the 5'-TTCC-3' (reverse) strand very weakly but detectably in our titration experiment. *A*, Replicate competitive titrations (in addition to **Figure 1B** in the main text) of PU.1ΔN167 with reverse hemi-methylated SC1. A reduction in fluorescence anisotropy was reproducibly observed as competitor concentration approaches 1 μM. To estimate the dissociation constant for the competitor, the unbound anisotropy was fixed at 0.308, an average value based on other titrations that reached complete displacement. *B* and *C*, This behavior was contrasted with the titration of Ets-1ΔN331 and auto-inhibited Ets-1ΔN280 with forward hemi-methylated SC1 under identical solution conditions. In this case, no systematic change in anisotropy was observed even at the highest competitor concentration (1 μM).

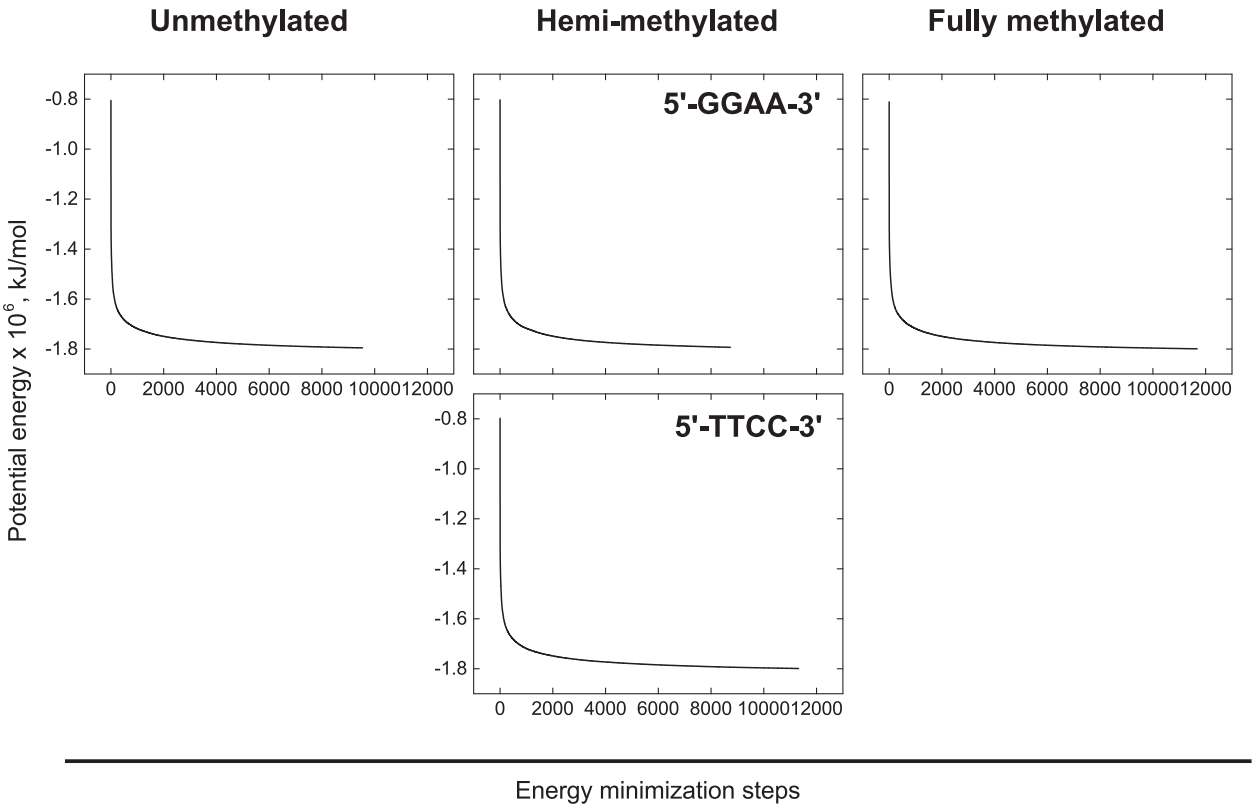

FIGURE S4. **Convergence of energy-minimized structures of un-, hemi-, and fully methylated DNA.** The total computed potential each structure was tracked at each step of the all-atom energy minimization procedure by steepest descent. Iteration was terminated when the maximum force could no longer be reduced, occurring for all structures at maximum forces below 1000 kJ mol^{-1} nm^{-1}.

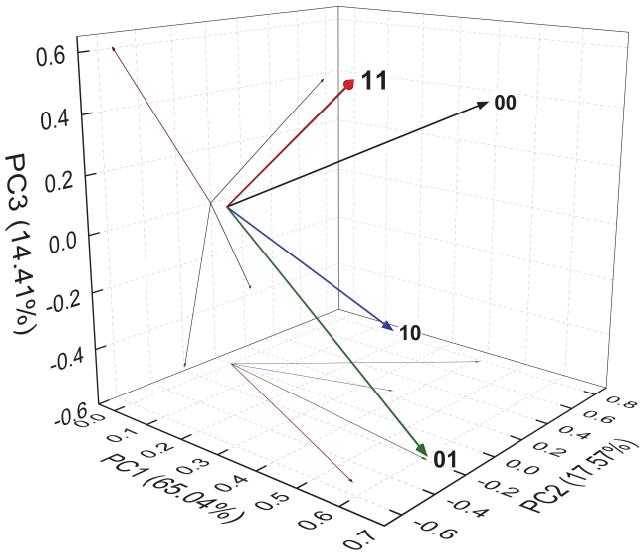

FIGURE S5. **Resolution of correlation in the reverse (5'-GGAA-3') glycosidic torsional angle χ between fully and hemi-methylated SC1 DNA.** This parameter shows correlation between fully methylated SC1 [11] and reverse hemi-methylated SC1 [01] when only PC1 and PC2 are considered (PC1-PC2 projection, equivalent to **Table 2** in the main text). Introduction of PC3, which is statistically different by Bartlett's test ($p < 0.05$), resolves this correlation, as the vectors 11 and 01 are split into opposite signs by PC3.

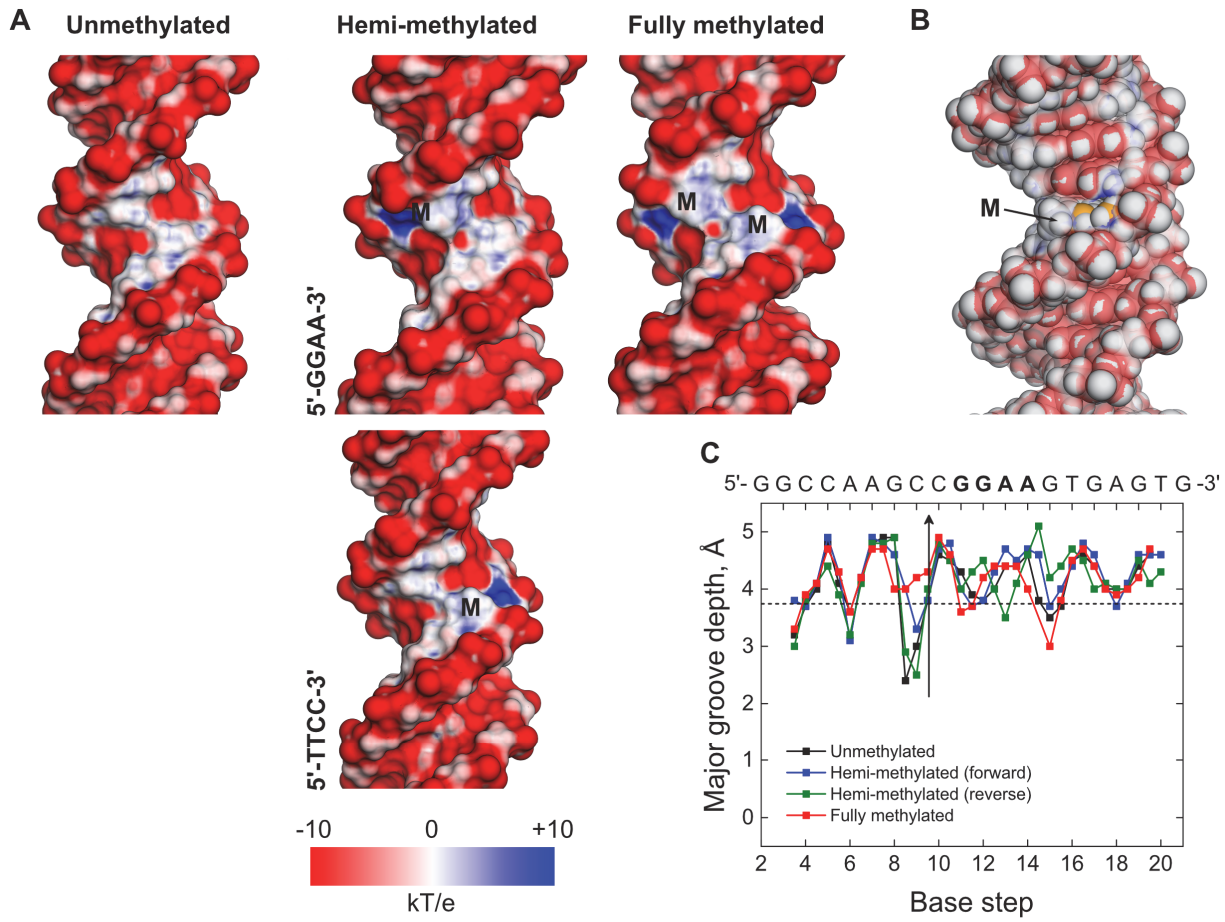

FIGURE S6. **Computed electrostatic properties of un-, hemi-, and fully methylated SC1.** *A*, Surface van der Waals potential of each energy-minimized structure was calculated according to continuum Poisson-Boltzmann electrostatics at physiologic ionic strength, as described in *Supplemental Methods*. The viewer is facing the major groove centered at the CpG dinucleotide of the SC1 site. The position of the 5-methyl substituent is marked as "M." *B*, The patches of positive charge in the methylated structures arise from increased exposure of the C6 and C2' of the 5-methylcytosines, as shown for the reverse hemi-methylated (5'-TTCC-3') sequence. The view follows a clockwise 90° axial turn in the plane of the page (major groove now facing left). For emphasis, the surface potential associated with C6 and C2' (yellow atoms) and their attached hydrogens (white) is removed. *C*, Major groove depth along the DNA sequence. The groove depth is locally increased by 0.5 Å (marked by arrow), without any corresponding change in the major groove width (not shown). The dashed line marks the average depth of the pre-minimized sequence with fiber geometry.

SUPPLEMENTAL REFERENCES

1. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, **31**, 5108-5121.

2. Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. *et al.* (2009) CHARMM: The biomolecular simulation program. *J Comput Chem*, **30**, 1545-1614.

3. Wang, P., Brank, A.S., Banavali, N.K., Nicklaus, M.C., Marquez, V.E., Christman, J.K. and MacKerell, A.D. (2000) Use of Oligodeoxyribonucleotides with Conformationally Constrained Abasic Sugar Targets To Probe the Mechanism of Base Flipping by HhaI DNA (Cytosine C5)-methyltransferase. *J Amer Chem Soc*, **122**, 12422-12434.

4. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*, **37**, 5917-5929.

5. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, **98**, 10037-10041.

6. Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G. and Baker, N.A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*, **35**, W522-W525.

7. Xiao, Q., Zhang, F., Nacev, B.A., Liu, J.O. and Pei, D. (2010) Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry*, **49**, 5588-5599.

8. Stephens, D.C., Kim, H.M., Kumar, A., Farahat, A.A., Boykin, D.W. and Poon, G.M. (2016) Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis. *Nucleic Acids Res*. doi: 10.1093/nar/gkw229.