Supporting Information

Lazarovici et al. 10.1073/pnas.1216822110

SI Methods

Cell Culture and DNA Extraction. IMR90 human fetal pulmonary fibroblast cells (ATCC) were cultured in a 5% (vol/vol) CO_2 humidified incubator. Cells were passaged to 70% confluence, and harvested using 15 mL Accutase. Cell viability was confirmed using Trypan blue staining. DNA was extracted from 5×10^6 cells using a 1:1 mixture of phenol–chloroform (phase lock, Eppendorf), and cleaned and concentrated using a minielute column (Qiagen).

DNase I Treatment of Purified DNA. Ten micrograms of purified, deproteinized DNA was suspended in 160 μL DNase I buffer (15 mM Tris·Cl, pH 8.0, 88.5 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM Spermidine, 6 mM $CaCl_2$) by aliquotting 20 μL into each of eight PCR tubes. An equal volume of DNase I buffer with DNase I was added to the DNA (0, 50, 25, 10, 5, 3, 1, and 0.5 units/mL, respectively) and incubated for 3 min at 37 °C, after which 10 μL of 75 mM EDTA was added to the tubes to terminate the enzymatic reactions. One microliter of digested material from each treatment was loaded on a 2% TAE agarose gel and run for 1 h at 100 V. The gel was stained with SYBR green and imaged on a Typhoon imager (Amersham) to assess the degree of digestion. Treatments with moderate levels of digestion (average fragment size >400 bp) were pooled. The pooled sample was layered onto a sucrose gradient containing 9.5 mL each of sucrose concentration of 40%, 30%, 20%, and 10% (wt/vol) containing 1 M NaCl, 10 mM Tris·HCl, pH 8.0, 5 mM EDTA, and ultracentrifuged for 24 h at 25,000 rpm at 16 °C in a SW21 swinging bucket rotor Beckman LE-80 Ultracentrifuge 77,002 g. We took 0.5 mL fractions from the top down, and fractions with fragments below 500 bp were pooled and cleaned up on a Qiagen PCR column.

Digital DNase I Mapping in IMR90 Cells. Digital DNase I mapping was performed essentially as described in ref. 1. Briefly, IMR90 cells were grown as described above. We pelleted 1×10^8 cells and washed them with cold PBS. Cell pellets were resuspended in Buffer A (15 mM Tris·Cl, pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0, 0.5 mM spermidine, and 0.15 mM spermine) to a final concentration of 2×10^6 cells/mL. Nuclei were obtained by dropwise addition of an equal volume of Buffer A containing 0.04% Nonidet P-40 to the cells, followed by incubation on ice for 10 min. Nuclei were centrifuged at 1,000 *g* for 5 min and then resuspended and washed with 25 mL of cold Buffer A. Nuclei were resuspended in 2 mL of Buffer A at a final concentration of 1×10^7 nuclei/mL. We performed DNase I (Roche, 60 U/mL) digests for 3 min at 37 °C in 2 mL volumes of DNase I buffer (13.5 mM Tris·HCl, pH 8.0, 87 mM NaCl, 54 mM KCl, 6 mM CaCl2, 0.9 mM EDTA, 0.45 mM EGTA, 0.45 mM Spermidine). Reactions were terminated by adding an equal volume (2 mL) of stop buffer (1 M Tris·Cl, pH 8.0, 5 M NaCl, 20% SDS, and 0.5 M EDTA, pH 8.0, 10 μg/mL RNase A, Roche) and incubated at 55 °C. After 15 min, we added Proteinase K (25 μg/mL final concentration) to each digest reaction and incubated for 1 h at 55 °C. After DNase I treatments, careful phenol-chloroform extractions were performed. Control (untreated) samples were processed as above except for the omission of DNase I. DNase I double-cut fragments were purified and sequencing libraries were constructed as described in ref. 2.

Library Construction and Sequencing. Libraries were constructed following Illumina's protocol, except for a few minor modifications. Sub–500 bp DNA fragments were combined with 1× T4 DNA ligase buffer (0.1 mM ATP) (NEB, New England Biolabs), 0.4 mM dNTP mix, 5 U *Escherichia coli* DNA polymerase I Klenow fragment (NEB), and 50 U T4 polynucleotide kinase (NEB). To repair blunt ends and end-phosphorylate, the reaction was incubated at 20 °C for 30 min and DNA was recovered using the MinElute microspin column (Qiagen Inc.). The repaired DNA fragments were subjected to adenylation by adding nontemplated adenines to the 3′ ends using 1 mM dATP (Sigma) and 5 U Klenow fragment (3′–5′ exo-) (NEB) and incubated at 37 °C for 30 min. After eluting the adenylated DNA through a MinElute column, Illumina adapters were ligated to the ends of the fragments in 50 μL reactions comprised of 25 μL 2× Quick DNA ligase buffer (NEB), 17.5 μL adenylated DNA fragments, 2.5 μL adapter oligo mix (Illumina), and 5 μL Quick DNA ligase (NEB). Ligated fragments were eluted from a MinElute column (Qiagen) after 15 min incubation at 20 °C. The adapter-ligated DNA was PCR enriched in a 50 μL reaction containing 15 μL adapter-ligated DNA, 25 μL 2× Phusion High-Fidelity PCR Master Mix (NEB), 8 μL of nuclease-free water (Ambion), and the addition of 1 μL of each Illumina primer 1.1 and 1.2 (Illumina). Several reactions were performed in parallel under the following thermal cycle profile: 98 °C for 30 s, 12 cycles of 98 °C for 15 s, 65 °C for 30 s, and 72 °C for 30 s, followed by 72 °C with an extension of 5 min. Enriched libraries were quantified using a Qubit (Invitrogen) after purification with a MiniElute microspin column (Qiagen). Libraries were sequenced on an Illumina GAIIx sequencer to obtain 36 bp single-end reads. The 5′ end of each read corresponds to a nucleotide immediately 3′ of a cleaved phosphate. In our analyses, information about the nucleotides 5′ of the cleaved phosphates was obtained by aligning quality-filtered raw reads to the human genome sequence using ELAND.

Single-Nucleotide Model. To determine over what spatial range cleavage rate is influenced by base identity at neighboring nucleotides, we counted how often base b (taking values A, C, G, and T) was found at nucleotide position j (where −1 denotes the first upstream nucleotide, and +1 the first downstream nucleotide, etc.) at uniquely mappable positions, and denoted this number by N_{jb}. We also determined a corresponding count K_{jb} for the set of all uniquely mappable cleavage events. Since the ratios $\lambda_{jb} = K_{jb}/N_{jb}$ scale with the total number of cleavage events, which is not relevant to the intrinsic properties of DNase I, we obtained a normalized relative cleavage rate parameter α_{jb} (taking values between 0 and 1) by dividing by the largest λ_{jb} for each j. To predict the relative cleavage rate for a given hexamer under the independent-nucleotide assumption, we computed the product of single-nucleotide cleavage rates over nucleotide positions:

$$\alpha_{\text{indep}}(h) = \prod_{j=-3}^{3} \sum_{b} \alpha_{jb} I_{jb}(h),$$

where the sum over b = A,C,G,T serves to select the appropriate α_{jb} at each position, the indicator function $I_{jb}(h)$ equaling unity if the base identity at position j in hexamer h is b, and zero otherwise.

Hexamer Model. Each hexamer was considered to be centered on the cleaved phosphate. All our models were constructed in a strand-specific manner. Only uniquely mapping reads were considered in the analyses. To construct a hexamer-based model of relative cleavage rate, we performed a similar procedure as above, counting for each hexamer context h the number N_h of mappable genomic positions and the number K_h of cleavage events. In this case, assuming that the hexamer fully specifies the relevant sequence context of each potential cleavage site in the genome, the ratio $\lambda_h = K_h/N_h$ can be interpreted as the expectation value of the number of cleavage events for an individual site of type h, where the latter follows a Poisson distribution. A normalized cleavage rate α_h was obtained by dividing λ_h by the maximum value across all hexamers. The resulting hexamer table is available in Dataset S1.

Outlier Removal. Anomalous phosphate positions were excluded in every analysis. To this end, an initial table of expected cleavage counts λ_h was constructed as described above. Subsequently, for each individual phosphate, a P value was computed based on the assumption that the number of cleavages at that position follows a Poisson distribution with mean λ_h, where h is the sequence context of the phosphate. If the P value was smaller than 10^{-5}, the phosphate position was deemed to be anomalous and excluded from further analysis. New λ_h values were then computed based on the reduced universe of phosphates. This procedure was repeated until convergence.

Estimating the Contribution of Sequencing Bias. To test whether our observed $\sim 10^3$ fold variation in DNase I cleavage rates could have resulted from sequencing bias, we analyzed a dataset prepared by sonicating purified genomic DNA. This dataset served as a control in the paper that introduced the Sono-Seq method (3). We found only a 30-fold variation in breakage rates at the hexamer level when using our model to analyze this control dataset. We therefore sought to determine whether this variation was attributable to sequencing bias or to preferential breakage intrinsic to DNA. The DNase I cleavage rates and sonication breakage rates do not appear to share any common biases. For the sonication control, phosphates with hexamer sequence context NNCpGNN broke at the highest rate. By contrast, we observed that DNase I preferentially cleaves phosphates with sequence context NNTpTNN. Not surprisingly, we found only a very weak correlation ($R^2 = 0.02$) between the hexamer cleavage rates obtained from the DNase I and sonication datasets. To further examine the degree to which the dataset from Auerbach et al. (3) captures preferential breakage intrinsic to DNA, we modified our algorithm and constructed a dinucleotide table, as opposed to a hexamer table. We compared these 16 dinucleotide breakage rates to those obtained in a separate study on sequence-specific ultrasonic cleavage of DNA (4). These authors constructed a dinucleotide table by collecting fragments of sonicated DNA and analyzing them using gel electrophoresis. We obtained good agreement with the dinucleotide breakage rates that Grokhovsky et al. (4) list in Table 1 of their paper ($R^2 = 0.61$). This suggests that rates obtained using our algorithm do not need to be adjusted to account for sequencing biases. The good ($R^2 = 0.54$) agreement of our DNase I cleavage rates with Herrara and Chaires (5), who used a completely different assay not subject to sequencing bias, makes it even more unlikely that our data are the results of sequencing bias.

Dependency Between Mutated and Modulating Positions. To systematically map nucleotide dependencies for all combinations of mutated and modulating bases within the hexamer context, we defined the "unconditional" free-energy change $\Delta\Delta G(b_m \rightarrow b'_m)$ in units of RT as the negative of the logarithm of the fold change in cleavage rate when the base at mutated nucleotide position m is mutated from b to b', regardless of sequence context. We defined the "conditional" free-energy change $\Delta\Delta G(b_m \rightarrow b'_m / b''_c)$ as the same quantity conditional on having base b'' at context-nucleotide position c.

Statistical Significance of Nucleotide Dependencies. To assess the statistical significance of the dependency between any pair of nucleotide positions within the hexamer, we first determined $K_{jbj'b'}$, the number of mappable cleavage events for which the base identities at nucleotide positions j and j' are b and b', respectively. Next, to obtain a (two-sided) P value, we used a null model where the nucleotides were considered to contribute independently to the cleavage rate and in which $K_{jbj'b'}$ follows a binomial distribution, with a sample size equal to the total number of unique mappable reads, and a "success probability" given by:

$$p_{jbj'b'} = \frac{N_{jbj'b'}\lambda_{jb}\lambda_{j'b'}}{\sum_{b'}\sum_{b''} N_{jb'j'b''}\lambda_{jb'}\lambda_{j'b''}}.$$

Here $N_{jbj'b'}$ equals the number of unique mappable genomic positions of a given type.

High-Throughput Prediction of Minor Groove Width. Minor groove width (MGW) was predicted for each of the 4,096 unique hexamers with a high-throughput (HT) approach. This method employs a sliding pentamer window in data mining 2,121 Monte Carlo (MC) trajectories for DNA sequences of 12–27 bp in length (Fig. 2*A*). The MC approach and protocol are described elsewhere (6–8). In the HT approach, the central base pair of each pentamer was assigned a MGW value equal to the average of all occurrences of that pentamer in our MC dataset. We validated this HT method based on a comparison of our MGW predictions with all crystal structures of protein–DNA complexes in the Protein Data Bank of at least one helical turn (10 bp) in length and without drastic deformations or chemical modifications. The Spearman's rank correlation coefficient between our MGW predictions and experimental data was 0.67. To evaluate the impact of crystal packing and protein binding, we also compared our MGW predictions for solution-state DNA structures for the Dickerson dodecamer derived from NMR spectroscopy using residual dipolar coupling (9), which yields excellent quantitative agreement (Pearson correlation coefficient of 0.84). Note that we previously used a comparable HT approach based on a combination of tetramers and pentamers (10). For a given hexamer $N_{-3}N_{-2}N_{-1}|N_{+1}N_{+2}N_{+3}$, the MGW predictions for positions N_{-1} and N_{+1} were unambiguous. For the remaining positions, we predicted MGW as the average over all 16 possible dinucleotide flanks.

MC Prediction of DNA Structures for Unmethylated and Methylated DNA. To predict MGW and roll of unbound DNA targets, we performed all-atom MC simulations for all 256 possible hexamers of the form NNNCGN, and the 256 corresponding sequences with 5-methylcytosines at both base pairs of the CpG dinucleotide at positions +1 and +2 (Figs. S7 and S8). We performed one MC simulation each for the unmethylated and methylated forms of these 256 hexamers after adding a CGCG flank on both sides. To evaluate end-effects for the hexamer ACTCGA, whose DNase I cleavage rate increased ~eightfold upon methylation, we performed nine independent MC simulations for this sequence (Fig. S9), three for each of the three different tetramer flanks (CGCG-N_6-CGCG, CGTA-N_6-TACG, and CATG-N_6-CATG). These runs were performed for the unmethylated and CpG methylated variants of the ACT|CGA hexamer. MC predictions yield ensembles of all-atom structures, which we analyzed for structural changes upon methylation. MGW and roll showed most significant effects upon CpG methylation. We used the CURVES algorithm (11) for deriving average MGW and roll from MC ensembles of 150,000 structures. For the ACTCGA sequence, we computed MGW and roll as the median over the predictions with the three different flanking sequences (compare with Fig. 4*C*). The MC methodology was previously described and validated (6, 8, 12, 13). The sampling methodology was now expanded by facilitating one additional internal degree of freedom, the rotation of the cytosine 5-methyl group, which we implemented in analogy to the rotation of the thymine methyl group (6, 7). Partial charges for 5-methylcytosine were taken from a set of AMBER force field parameters derived for chemically modified nucleotides (14).

Dependence of DNase I Cleavage Rate on DNA Methylation Status. Virtually all methylation in IMR90 occurs within CpG dinucleotides (15). We used the genome-wide methylation map for the IMR90 cell line to define two classes of phosphate positions as follows. For the "high-methylation" class, we first determined which CpG dinucleotides were in the bottom 80% in terms of degree of methylation, and then excluded any phosphate whose hexamer context overlapped with one or more such CpGs by at least one base on either strand. For the "low-methylation" class, we similarly excluded phosphates overlapping with any CpGs in the top 80%. Hexamer-based cleavage rate models were then constructed for each class separately (compare with Fig. 3*A*). These are available in Dataset S1. Another goal was to systematically discover sequence features whose presence within the hexamer context causes the cleavage by DNase I to be affected by DNA methylation status. We considered two types of features: (*i*) single-nucleotide features of type $f = (j,b)$ spanning all 6×4 possible combinations of position j and base b, and (*ii*) all 140 dinucleotide features of type $f = (j,j',b,b')$. For each feature, we computed $\lambda_f = K_f/N_f$ for each class, and then took the ratio of the high- and low-methylation value. A significant deviation of this ratio from unity indicated a dependency on methylation status (compare with Fig. S6).

Inferring Methylation Status from in Vivo DNase I Footprints. We tested our ability to predict DNA methylation status using DNase I cleavage data for the IMR90 cell line. To have sufficient statistical power, we limited our analysis to a set of 455 nonoverlapping windows of size 2,500 bp in which at least 400 cleavages adjacent to a CpG dinucleotide were observed. For each window, we observed the number K_{CG} of cuts adjacent to a CpG dinucleotide within the window. To compute the expected number of such cuts, we started from the table of relative cleavage rates for hexamers derived from genome-wide DNase I data for DNA purified from IMR90 cells as described above. The probability f_{CG} that a cleavage event within the window occurs adjacent to a CpG dinucleotide equals the sum of cut rates over all hexamers of type NNNCGN divided by the sum over all hexamers within the window on either strand. To obtain the expected value of K_{CG}, we multiplied f_{CG} by the total number N of reads within the window. Since we found that methylated CpGs are cleaved at a higher rate (compare Fig. 2*A*), values of $K_{CG}/(f_{CG}*N)$ significantly higher/lower than unity predict regional hyper/hypomethylation, respectively.

1. Thurman RE, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82.
2. Sabo PJ, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3(7):511–518.
3. Auerbach RK, et al. (2009) Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci USA* 106(35):14926–14931.
4. Grokhovsky SL, et al. (2011) Sequence-specific ultrasonic cleavage of DNA. *Biophys J* 100(1):117–125.
5. Herrera JE, Chaires JB (1994) Characterization of preferred deoxyribonuclease I cleavage sites. *J Mol Biol* 236(2):405–411.
6. Rohs R, Sklenar H, Shakked Z (2005) Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites. *Structure* 13(10):1499–1509.
7. Rohs R, Bloch I, Sklenar H, Shakked Z (2005) Molecular flexibility in ab initio drug docking to DNA: Binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic Acids Res* 33(22):7048–7057.
8. Joshi R, et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131(3):530–543.
9. Wu Z, Delaglio F, Tjandra N, Zhurkin VB, Bax A (2003) Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy. *J Biomol NMR* 26(4):297–315.
10. Slattery M, et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147(6):1270–1282.
11. Lavery R, Sklenar H (1989) Defining the structure of irregular nucleic acids: Conventions and principles. *J Biomol Struct Dyn* 6(4):655–667.
12. Rohs R, West SM, Liu P, Honig B (2009) Nuance in the double-helix and its role in protein-DNA recognition. *Curr Opin Struct Biol* 19(2):171–177.
13. Bishop EP, et al. (2011) A map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. *ACS Chem Biol* 6(12):1314–1320.
14. Aduri R, et al. (2007) AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory Comput* 3(4):1464–1475.
15. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322.

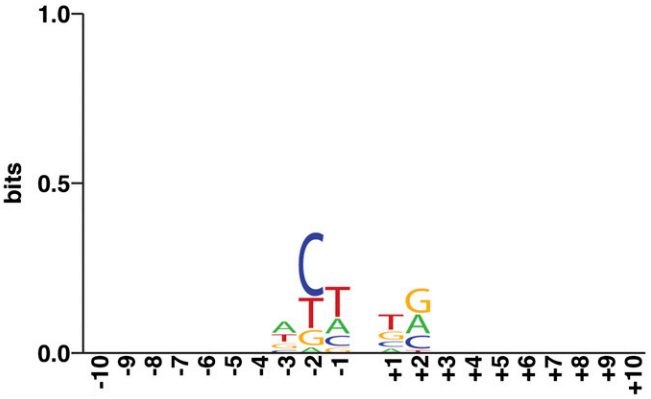

Fig. S1. The dependence of DNase I cleavage rate on base identity is largely limited to a hexamer window centered at the cleaved bond. We derived position-specific relative cleavage rates from a DNase I digestion of human genomic DNA under the assumption of independence between nucleotides. Shown is the information content at each nucleotide position relative to the cleaved phosphate bond, revealing that the dependence of the intrinsic DNase I cleavage rate on DNA sequence context is largely limited to the region from 3 nt upstream to 3 nt downstream.

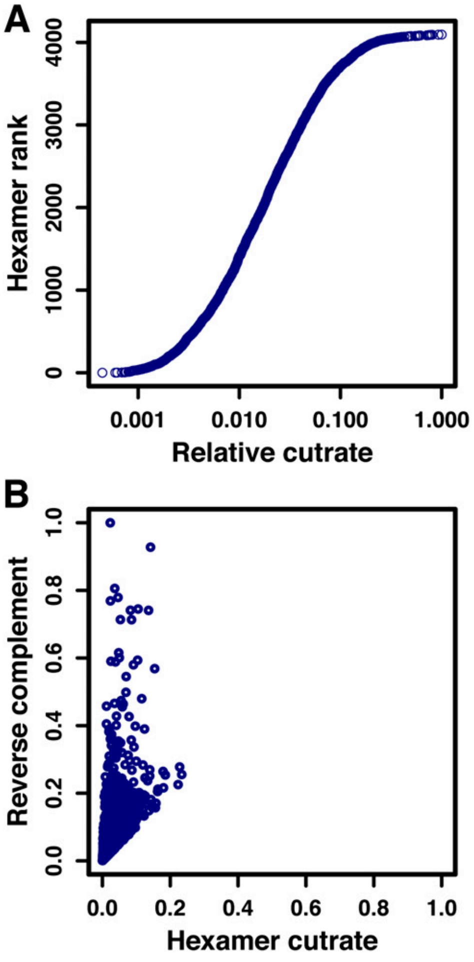

Fig. S2. Global properties of the hexamer-based model of intrinsic DNase I cleavage rate. (*A*) Cumulative distribution function of cleavage rate, showing that its variation with local sequence spans three orders of magnitude. (*B*) A direct comparison of the cleavage of hexamers and their reverse complements reveals that DNase I cuts DNA in a highly strand-specific manner. Palindromic hexamers were excluded from this plot.

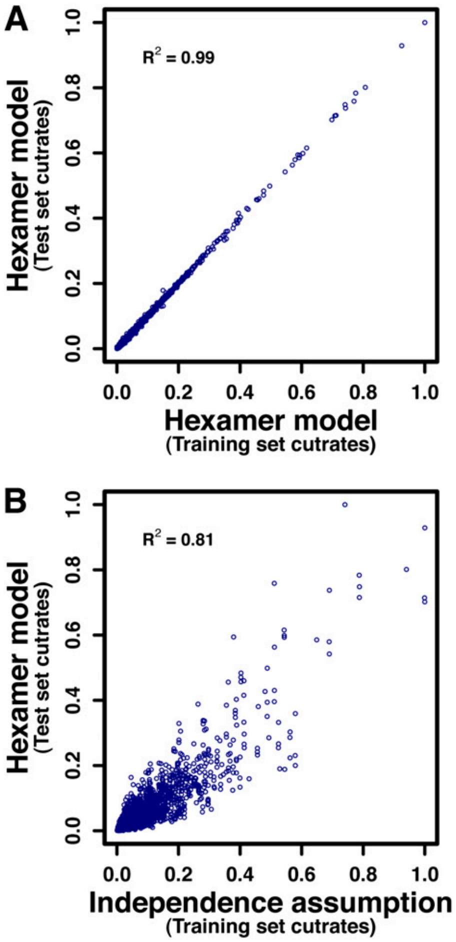

Fig. S3. Hexamer cleavage rates are highly reproducible and indicate the existence of dependencies between nucleotide positions. (*A*) Cleavage rates computed using the full hexamer model (see Hexamer model in *SI Methods*) for both the training and test set data. (*B*) Cleavage rates computed using the independence assumption (see "Single-Nucleotide Model" in *SI Methods*) for the training set data are compared with those obtained using the full hexamer model for the test set. Clearly the agreement is not as good when the independence assumption is used to compute cleavage rates.

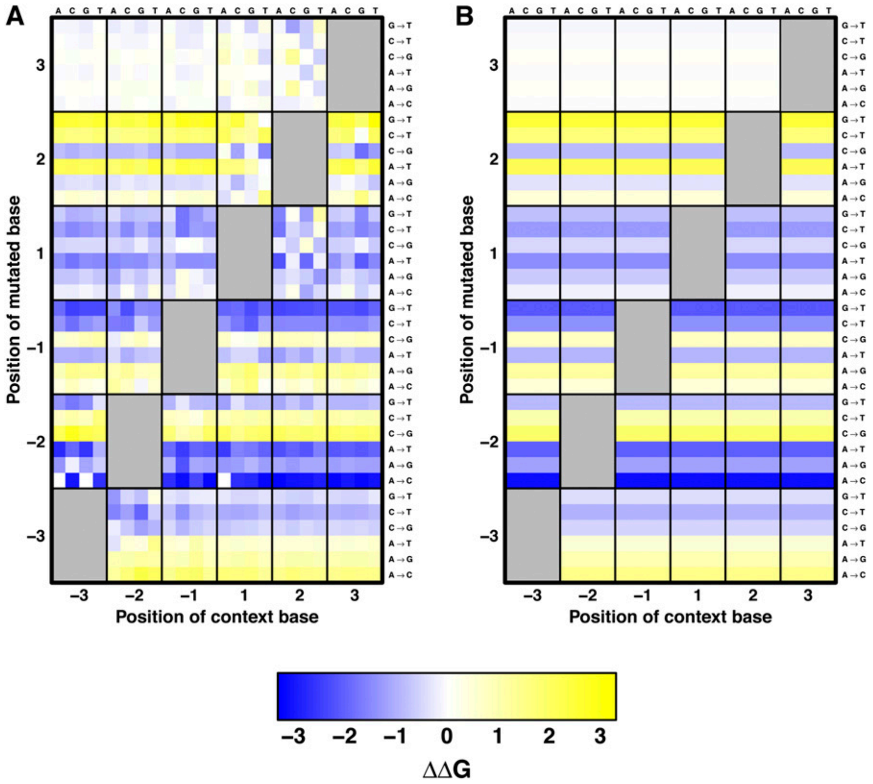

Fig. S4. Conditional and unconditional free energy parameters. (*A*) Heat map representation of the conditional ΔΔG for each possible combination of a mutation $b_m \to b_m'$ at nucleotide position *m* and base b_c'' at context position *c*. (*B*) The unconditional ΔΔG, which by definition is independent of the context nucleotide (i.e., uniform rows in heat map).

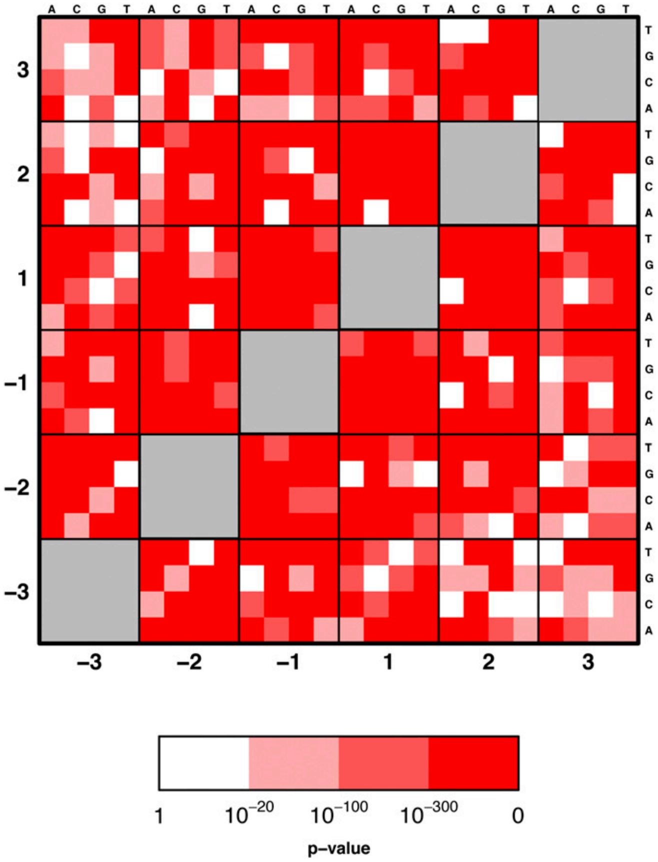

Fig. S5. Overview of dependencies within the local sequence context. Statistically significant dependencies between nucleotide positions can be detected throughout the hexamer window.

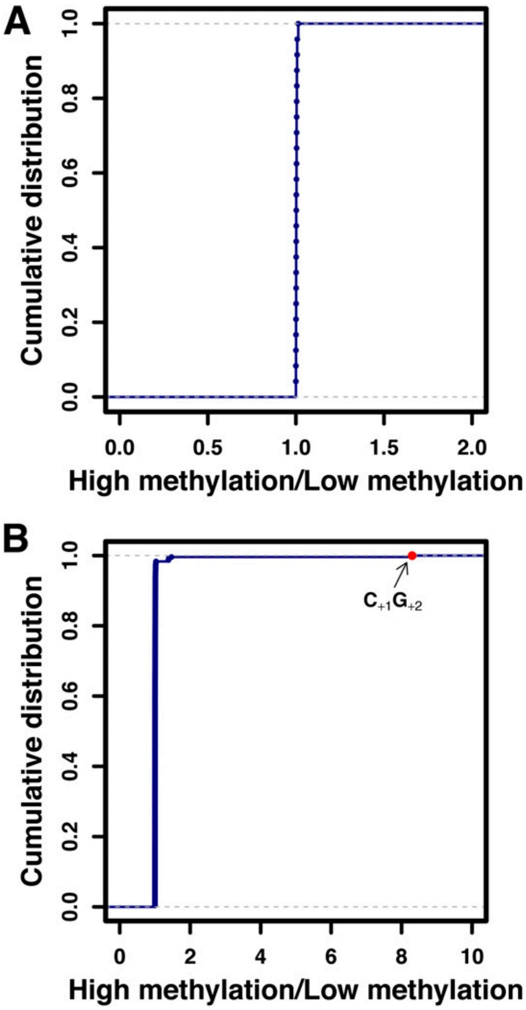

Fig. S6. Systematic search for sequence features underlying the methylation dependence of DNase I cleavage rate. Shown is the (cumulative) distribution of fold-differences associated with all possible (*A*) mono- and (*B*) dinucleotide-features within the hexamer context. The only feature that shows a significant (~eightfold) effect is a CpG dinucleotide immediately downstream of the cleaved phosphate.

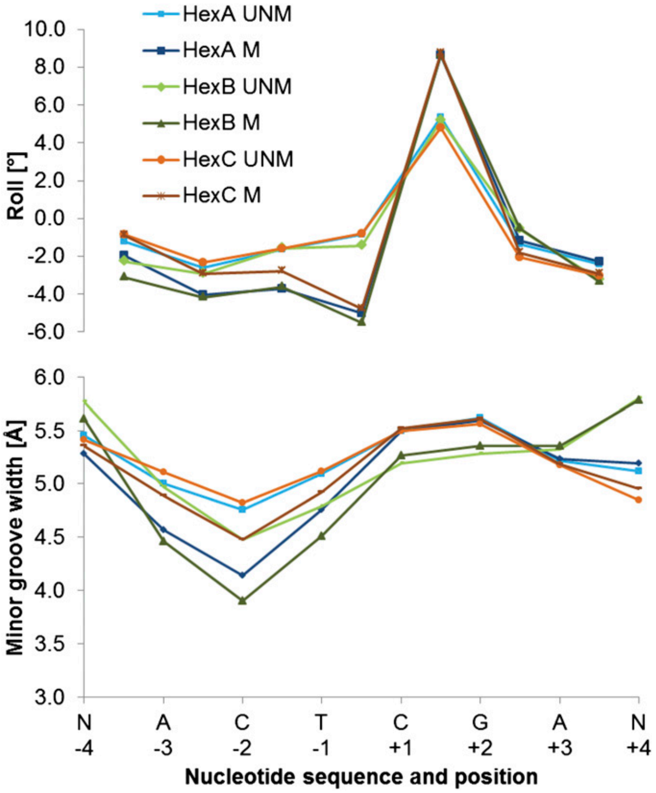

Fig. S7. Effect of CpG methylation on roll and MGW for different flanking sequences. MC predictions of roll and MGW for methylated (M) and unmethylated (UNM) versions of the same hexamer, averaged over three different flanking sequences, denoted by HexA (CGCG), HexB (CGTA) and HexC (CATG). Independent of flanking sequence, methylation leads to an increase in the positive roll angle at the CpG dinucleotide and a narrowing of the MGW at position −2 by roughly 0.5 Å. The prediction for each flanking sequence (*SI Methods*) is based on three independent MC runs.

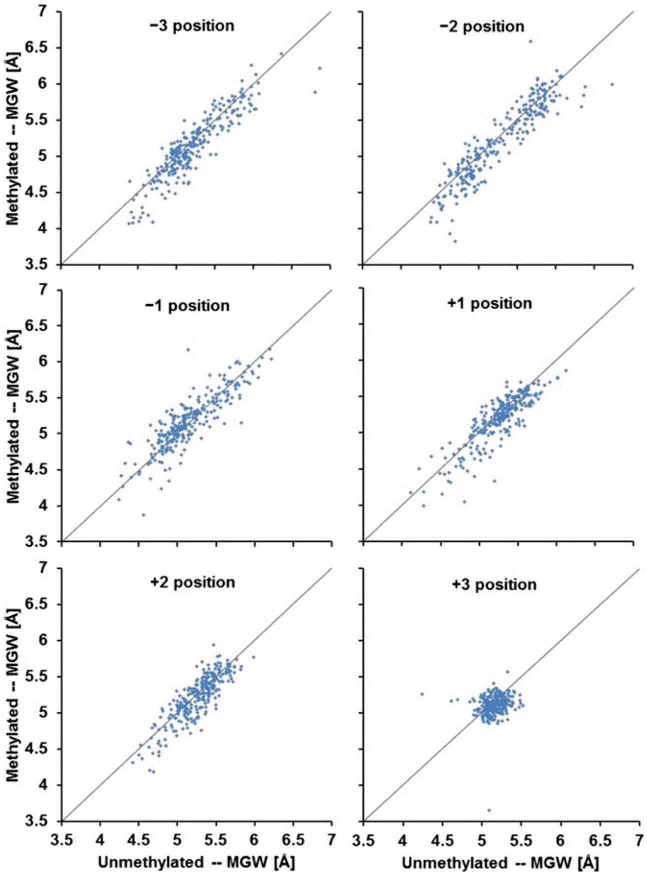

Fig. S8. Effect of CpG methylation on MGW. For each unique hexamer of type $N_{-3}N_{-2}N_{-1}C_{+1}G_{+2}N_{+3}$, MC simulations were used to predict MGW of all 256 sequences of the form $CGCGN_{-3}N_{-2}N_{-1}C_{+1}G_{+2}N_{+3}CGCG$ for both unmethylated and fully methylated states of the $C_{+1}G_{+2}$ dinucleotide. Consistent with the observation shown in Fig. 4C, for hexamers that already assume a narrow MGW at position −2 in their unmethylated state, $C_{+1}G_{+2}$ methylation tends to further narrow the groove at this position.

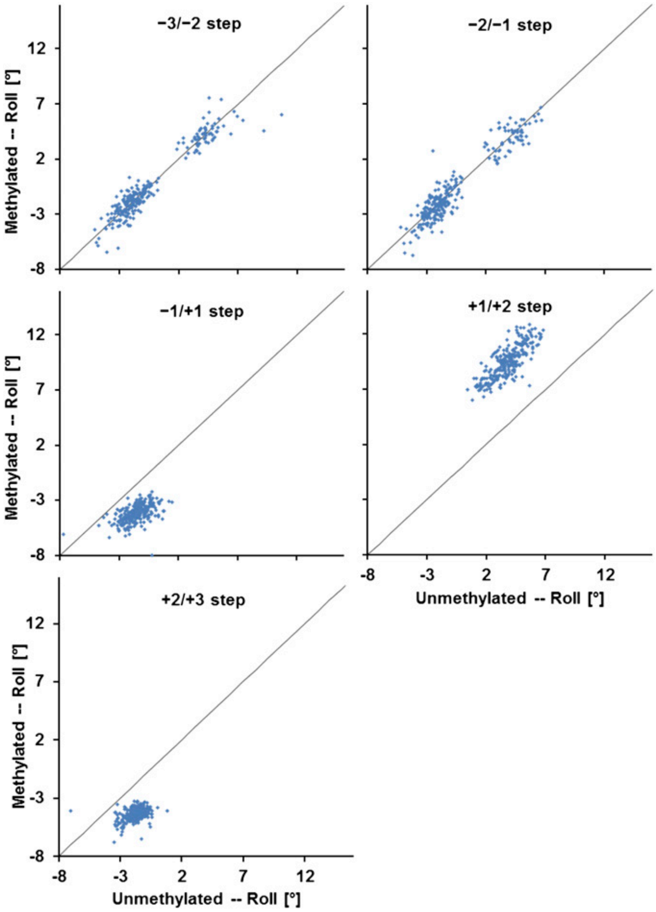

Fig. S9. Effect of CpG methylation on roll angles. Analyzing the same MC simulations described in Fig. S8, we analyzed the effect of $C_{+1}G_{+2}$ methylation on roll angles at all five dinucleotide steps within the 256 unique hexamers of the form $N_{-3}N_{-2}N_{-1}C_{+1}G_{+2}N_{+3}$. Our results show that $C_{+1}G_{+2}$ methylation monotonically increases roll angles at the +1/+2 base pair step for all hexamers. Such increase is partially absorbed by decreases in roll angles at the adjacent −1/+1 and +2/+3 base pair steps.

Table S1. Construction of position-specific cleavage rate matrix

Base	−3	−2	−1	+1	+2	+3
Observed cuts ($\times 10^6$)						
A	7.636	1.396	4.721	3.141	6.793	5.251
C	1.813	7.979	2.270	2.355	3.233	3.510
G	2.608	2.115	1.369	3.304	5.753	3.299
T	5.221	5.787	8.919	8.479	1.498	5.219
Genomic positions ($\times 10^9$)						
A	1.496	1.496	1.496	1.496	1.496	1.496
C	0.999	0.999	0.999	0.999	0.999	0.998
G	0.998	0.998	0.998	0.998	0.998	0.998
T	1.484	1.484	1.484	1.483	1.484	1.484
Relative cleavage rates						
A	1.000	0.117	0.525	0.367	0.786	0.998
C	0.356	1.000	0.378	0.414	0.561	1.000
G	0.511	0.266	0.229	0.580	1.000	0.939
T	0.689	0.489	1.000	1.000	0.175	0.999

For each subclass of phosphates determined by a combination of relative nucleotide position and base identity, the total number of cleavage events and the total number of mappable genomic positions were determined. The ratio of these counts was then scaled to a maximum of unity for each position, resulting in a matrix of relative cleavage rate parameters.

Dataset S1. Hexamer tables for DNase I cleavage rates

Dataset S1

Hexamer tables for DNase I cleavage rates were prepared using three different populations of phosphates: (First spreadsheet) All phosphates were used, regardless of methylation status of cytosines in hexamer window. (Second spreadsheet) Only phosphates whose hexamer sequence context exhibited a low degree of cytosine methylation were used. (Third spreadsheet) Only phosphates whose hexamer sequence context exhibited a high degree of cytosine methylation were used.