Published in final edited form as:

Biochemistry. 2020 May 26; 59(20): 1909–1926. doi:10.1021/acs.biochem.0c00253.

A conformational switch in the zinc finger protein Kaiso mediates differential readout of specific and methylated DNA sequences

Evgenia N. Nikolova[†], Robyn L. Stanfield, H. Jane Dyson, Peter E. Wright^{*}

Department of Integrative Structural and Computational Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037

Abstract

Recognition of the epigenetic mark 5-methylcytosine (mC) at CpG sites in DNA has emerged as a novel function of many eukaryotic transcription factors (TFs). It remains unclear why the sequence specificity of these TFs differs for CpG-methylated motifs and consensus motifs. Here, we dissect the structural and dynamic basis for this differential DNA-binding specificity in the human zinc finger TF Kaiso, which exhibits high affinity for two consecutive mCpG sites in variable contexts and also for a longer, sequence-specific Kaiso binding site (KBS). By integrating structural analysis and DNA binding studies with targeted protein mutagenesis and nucleotide substitutions, we identify distinct mechanisms for readout of methylated and KBS motifs by Kaiso. We show that a key glutamate residue (E535), critical for mCpG site recognition, adopts different conformations in complexes with specific and methylated DNA. These conformational differences, together with intrinsic variations in DNA flexibility and/or solvation at TpG versus mCpG sites, contribute to different DNA affinity and sequence specificity. With methylated DNA, multiple direct contacts between E535 and the 5' mCpG site dominate the binding affinity, allowing for tolerance of different flanking DNA sequences. With KBS, Kaiso employs E535 as part of an indirect screen of the 5' flanking sequence, relying on key tyrosine-DNA interactions to stabilize an optimal DNA conformation and select against non-cognate sites. These findings demonstrate how TFs use conformational adaptation and exploit variations in DNA flexibility to achieve distinct DNA readout outcomes and target a greater variety of regulatory and epigenetic sites than previously appreciated.

Graphical Abstract

^{*}Corresponding Author Peter E. Wright, wright@scripps.edu.

[†]Present Address: Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore MD 21218

Supporting Information

Figures and tables with crystallographic and NMR information.

Accession numbers

The coordinates for the X-ray crystal structures have been deposited in the RCSB Protein Data Bank (PDB) with accession numbers 6DF5 (Kaiso-KBS at pH 7.2), 6DF8 (Kaiso-KBS at pH 6.5), 6DFB (Kaiso(K539A)-KBS), 6DF9 (Kaiso(E535Q)-KBS at pH 7.2), 6DFA (Kaiso(E535A)-KBS at pH 7.2), 6DFC (Kaiso-KBS(5TCU/A) at pH 7.2), 6V8U (Kaiso-KBS(5TCG) at pH 5.0).

Introduction

Transcription factors (TFs) play a central role in many biological processes by ensuring correct patterns of gene expression. Breakdown in their normal function can cause numerous human diseases.^{1,2} To stimulate or inhibit transcription, TFs recognize specific DNA sequences that are encoded in the regulatory regions of genes.^{3–5} A growing number of eukaryotic TFs are also found to bind effectively to DNA that is methylated at CpG dinucleotide sites.^{6–8} Such methylation of DNA is part of the epigenetic marking of the genome that constitutes an alternative code for gene expression.^{9,10} Epigenetic modifications of DNA and the core histone proteins involved in DNA packaging work in a highly orchestrated manner with many chromatin factors to alter the architecture of chromatin and assure proper transcriptional control.^{10–12} Thus, the ability of many TFs to “read” not only specific DNA sequence motifs but also epigenetic DNA methylation increases their functional diversity and provides another essential layer of gene regulation.

DNA methylation is one of the most widely studied epigenetic modifications.^{13,14} Methylation of mammalian DNA by the enzymatic addition of a methyl group to the C5 position of cytosine results in formation of 5-methylcytosine.^{15,16} Methylation occurs mainly in the context of CpG sites, where the cytosines on both strands are methylated.¹⁷ CpG methylation is typically linked to formation of a repressive chromatin structure and the inactivation of transcription.^{18–21} In normal cells, methylation at CpG clusters in promoter regions (“CpG islands”¹⁵) is critical for gene silencing during developmental processes such as genomic imprinting²² and X-chromosome inactivation.²³ In cancer, however, abnormal

increase in methylation of CpG islands is linked to silencing of tumor suppressor genes and cancer progression.^{24–26} While CpG methylation in gene promoters is associated with repression of transcription, CpG methylation within gene bodies (introns and exons) is correlated with activation of transcription.^{27–29} The function of gene-body methylation is not well understood, but has been shown to prevent spurious RNA polymerase entry and aberrant transcription initiation.³⁰

The interplay between DNA methylation and transcription factors is complex and our knowledge of it is constantly evolving. On one hand, methylation of DNA can hinder binding of TFs to their target sites, directly or by increasing the compaction and rigidity of chromatin.^{16,19,31} Conversely, DNA methylation can create new binding sites for special TFs called methyl-CpG binding proteins (MBPs).^{32,33} Association of MBPs with methylated CpG sites can, in turn, have multiple outcomes, including directly blocking access of other regulatory proteins, recruiting chromatin modifiers to establish a repressive chromatin environment, and even diverting the same MBPs from binding to nearby functional sites.^{20,31,34–36}

Many potential MBPs have been identified and their interactions with methylated DNA characterized. Initially, recognition of methylated CpG (mCpG) sites was thought to be unique to members of the mCpG-binding domain (MBD) family, represented by MeCP2 and MBD1, MBD2, and MBD4 in mammals.³² Subsequent studies revealed that other mammalian TFs can bind specifically to mCpG sites, including members of the BTB/POZ C₂H₂ zinc finger (ZF) family (Kaiso (Zbtb33), Zbtb4, and Zbtb38),^{33,37} the basic leucine zipper (bZip) CCAAT/enhancer binding protein- α (CEBP α),³⁸ the zinc finger protein Zfp57 and its cofactor KAP1,³⁹ and several other zinc finger TFs (i.e. WT1, Egr1, KLF4, and CTCF).^{40,41} More recently, advanced high-throughput *in vitro* and *in vivo* technologies have uncovered a multitude of other TFs that preferentially recognize mCpG sites,^{7,42–44} suggesting that this phenomenon is far more widespread than previously thought. Currently-known MBPs mainly comprise zinc finger proteins and members of the extended homeodomain family (i.e. HOXB13, Oct4, etc.).^{6,7} A prevalent feature among MBPs is their function in embryonic development, where their activity could promote major changes in the state of chromatin. Notably, many of these MBPs display altered sequence specificity when binding to mCpG-containing motifs as compared to their known consensus motifs.^{6,8} For example, Kaiso can effectively bind two adjacent mCpG sites in variable sequence contexts but favors binding to a longer specific sequence, 5'-TCCTGCNA (N = any nucleotide).⁴⁵ Similarly, CEBP α binds a particular sequence motif, 5'-TGACGTCA, but loses specificity towards half of the motif with CpG methylation, 5'-mCGTCA.³⁸ Also, the canonical motif for KLF4, 5'-(A/G)(G/A)GGYGY (Y = T or C), is different from the preferred methylated motif, 5'-GGmCGGG.⁴² This implies that MBP recognition of mCpG sites may involve a distinct DNA binding mode or readout mechanism from that of specific motifs.

Methylation of CpG sites can modulate MBP binding through changes in the direct or indirect readout of DNA. Direct readout occurs through specific interactions between amino acid side chains and functional groups on DNA bases in the minor and major groove that contribute to favorable protein-DNA binding.^{46,47} Indirect readout refers to sequence-dependent variability of the DNA conformation and dynamics that can optimize the

geometry at the protein-DNA interface.^{46,48} MBPs, despite their structurally diverse DNA-binding domains, use common strategies to recognize mCpG sites.^{8,49,50} The MBD and C₂H₂ ZF domains, specifically, use a spatially conserved arginine to form hydrophobic interactions with the mC methyl group and bi-dentate hydrogen bonds with the 3' guanine base. Another common feature is the presence of direct or water-mediated CH---O type hydrogen bonds to the mC methyl groups, typically formed via the carboxyl group of a glutamate or aspartate residue. Most well-characterized MBPs exhibit nanomolar binding affinities for mCpG sites (< 200 nM),^{40,51–54} though some bind remarkably tightly to methylated DNA (0.2 nM).^{53,54} Moreover, their selectivity for methylated CpG sites varies substantially (from 2- to 100-fold) depending on the extent of methyl-specific interactions.^{40,51–54}

Methylation at CpG sites can also affect the indirect readout of DNA. Although the addition of a methyl group in the major groove introduces a subtle structural change,⁵⁵ it can impart large changes in the local and global physical properties of DNA.⁵⁶ CpG dinucleotides are particularly flexible^{57–61} and exhibit diverse conformations that are strongly context-dependent;^{62–64} they are frequently found in protein-DNA complexes that feature deformations^{65–67} and are readily incorporated into nucleosomes.^{68,69} Methylation at CpG sites, however, makes them more rigid and harder to bend,^{59,70–73} and alters the properties of adjacent DNA segments.^{74,75} DNA containing mCpG sites displays lower propensity to circularize or form stable nucleosomes *in vitro*^{74,76} and tends to be depleted in nucleosomes *in vivo*.⁷⁷ Thus, the change in DNA conformation and flexibility with CpG methylation could modulate how MBPs recognize different DNA motifs.

The BTB/POZ C₂H₂ zinc finger transcription factor Kaiso⁷⁸ (UniProt Q86T24) represents an ideal system to dissect the mechanism by which MBPs differentially recognize and discriminate between specific and methylated target motifs. Kaiso recognizes two consecutive methylated CpG sites⁴⁵ and also binds specifically to a longer non-methylated motif known as the Kaiso binding sequence (KBS, 5'-TCCTGCNA) that features a TpG site in place of the 5' mCpG site.⁴⁵ Using both its methylation-dependent and sequence-specific DNA-binding activities, Kaiso regulates a diverse set of genes linked to normal development and cancer.^{79–84} While the KBS motif is largely conserved, *in vivo* methylated DNA targets of Kaiso contain variable sequences flanking the mCpG sites.^{45,85,86} Despite high-resolution structural studies of Kaiso bound to DNA, it remains unclear how Kaiso selects for different sequences in methylated versus specific motifs. Current data show a similar binding mode of Kaiso for KBS and methylated DNA, including in the vicinity of the core TpG and mCpG sites.^{53,87} An optimally positioned glutamate residue (E535) forms direct interactions with the two cross-strand mCs in the form of canonical hydrogen bonds with the mC N4 and CH---O hydrogen bonds with the mC methyl groups.⁵³ Interactions between E535 and the 5' mCpG site are critical for the exceptionally high affinity and selectivity of Kaiso for methylated DNA.^{53,88} E535 interacts similarly with the core TpG site in KBS but requires protonation to form a hydrogen bond between the glutamate side chain carboxyl group and T O4.⁸⁷ Here, we use structural and biochemical tools combined with targeted protein mutagenesis and nucleotide substitutions to probe the molecular basis by which Kaiso recognizes KBS and methylated DNA sites embedded within different flanking sequences.

We find that Kaiso employs different readout mechanisms for specific and methylated motifs that are facilitated by differences in E535 conformations.

Materials and Methods

Sample preparation.

Uniformly ^{15}N - or ^{13}C , ^{15}N -labeled protein constructs comprising the optimal DNA-binding domain (zinc fingers 1, 2 and 3) of human Kaiso (amino acids 472–604) were expressed and purified as previously described.⁸⁷ Single point mutants (E535A, E535Q, K539A, Y536A and Y562A) were generated using standard methods for site-directed mutagenesis and prepared as above. Kaiso was refolded by dissolving lyophilized protein in 10 mM Tris pH 7.0, 1 mM TCEP (NMR buffer) to $\sim 0.2 - 0.4$ mM for 1 hour and diluting the sample with NMR buffer supplied with 3.3X molar equivalents of Zn^{2+} to a final concentration of ~ 20 μM . The protein concentration was quantified before and after refolding using the absorbance at 280 nm (NanoDrop, Thermo Scientific) and an extinction coefficient of 21890 $\text{M}^{-1}\text{cm}^{-1}$ for Kaiso Y536A and Y562A and 23380 $\text{M}^{-1}\text{cm}^{-1}$ for Kaiso wild-type and other mutants (ExPASy ProtParam tool). Refolded protein stocks were stored at 4 °C for up to 2 weeks. All DNA oligonucleotides (Table S1), including 5-methylcytosine and 5'-biotin labeled oligonucleotides were obtained from IDT (Coralville, USA) with standard desalting. Oligos were dissolved in NMR buffer to 0.1–0.5 mM and quantified using absorbance at 260 nm with extinction coefficients provided by the manufacturer. To form duplex DNA, complementary strands were annealed in equimolar ratios by heating to 95 °C for 2 min and cooling at 25 °C for 30 min. Kaiso-DNA complexes for NMR or crystallization studies were formed by adding refolded protein to DNA in a 1:1.1 ratio, followed by extensive desalting and buffer-exchange against NMR buffer using an Amicon Ultra centrifugal filter unit with a 3 kDa cutoff (EMD Millipore). Final sample concentrations were $\sim 0.1 - 0.4$ mM for NMR and $\sim 0.4 - 0.6$ mM for crystallization. For NMR, complex formation was verified by monitoring 1D imino proton spectra for the DNA and 2D ^1H - ^{15}N HSQC spectra for the protein.

Biolayer interferometry.

DNA binding studies using biolayer interferometry were performed on an Octet RED96 system with Streptavidin (SA) biosensors (Pall ForteBio). DNA duplexes containing a 5'-biotin label on one strand were used for immobilization to the sensors (Table S1). The reactions were performed and optimized as previously described,⁵³ using 10 mM Tris (pH 7.0), 150 mM KCl, 1 mM TCEP, 0.05% Tween 20, 1 mg/ml dried milk powder as binding buffer and Kaiso concentrations in the range of 12.5 nM to 800 nM. The kinetic profiles were typically fitted using a two-state binding model incorporated in the Octet software to obtain association (k_{on}) and dissociation (k_{off}) rate constants, which were used to calculate the binding affinity (K_{d}). In situations where protein-DNA binding was weak or there were significant non-specific binding effects such that k_{on} could not be reliably fit, the binding profiles were analyzed as previously described.⁵³

X-ray crystallography.

Crystallization trials for the Kaiso-KBS complex were performed using the automated Rigaku CrystalMation system of the Joint Center for Structural Genomics (JCSG) at TSRI. Several conditions were replicated on a large scale using the sitting drop vapor diffusion method. Crystallization trays were set up by mixing 1 μ l of protein-DNA complex in NMR buffer with 1 μ l of crystallization buffer in a drop and using 0.5 to 1 ml of crystallization buffer for the reservoir. Trays were incubated at 20 °C or 4 °C and crystals were typically obtained within 3 days. The crystallization conditions for Kaiso-KBS, Kaiso(E535A)-KBS, Kaiso(E535Q)-KBS, Kaiso(K539A)-KBS, and Kaiso-KBS(5TCU/A) complexes were 0.2 M Di-sodium tartrate (pH ~ 7.2), 20% PEG3350 at 20 °C. For the Kaiso-KBS complex at pH 6.5, the pH of the crystallization buffer above was adjusted to 6.5 using 1M HCl. For the Kaiso-KBS(5TCG) complex, the crystallization conditions were 0.02 M calcium chloride, 0.1 M sodium acetate pH 5.0, 30 v/v 2-methyl-2,4-pentanediol (MPD) at 4 °C. Crystals were briefly immersed into crystallization buffer in the presence of 25% glycerol prior to cryo-cooling. Crystallographic data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL), Advanced Photon Source (APS) at Argonne National Laboratory, and the Advanced Light Source at Lawrence Berkeley National Laboratory (see Table S2). The data were processed using HKL-2000.⁸⁹ All structures were solved by molecular replacement with Phaser⁹⁰ using the previously published Kaiso-KBS structure (PDB: 4F6M)⁸⁷ as a search model. Multiple cycles of manual model building (Coot⁹¹) and refinement (Phenix⁹²) were completed. Distance and angle restraints were applied for Zn²⁺-Cys₂His₂ coordination. Hydrogen-bonding distance restraints were imposed on the three terminal DNA base pairs due to poor electron density. Structural figures were rendered using PyMOL⁹³ (Schrödinger, LLC). Detailed analysis of DNA structural parameters was performed using 3DNA 2.0.⁹⁴

NMR spectroscopy.

NMR spectra were acquired at 298 K on a Bruker Avance 800 MHz spectrometer equipped with a triple-resonance gradient cryo-probe or Bruker 500 MHz and 900 MHz spectrometers equipped with room-temperature probes. All spectra were processed and analyzed using NMRPipe⁹⁵ and Sparky.⁹⁶ Standard ¹H-¹⁵N HSQC spectra of Kaiso-DNA complexes (0.1–0.4 mM) in NMR buffer (5% D₂O) were acquired on each spectrometer. Backbone resonance assignments for the KBS complexes of the E535A, E535Q, Y536A, and K539A mutants were deduced from the published assignments of the wild type Kaiso-KBS complex.⁸⁷

The pH titration of the Kaiso-KBS complex was performed on a 0.15 mM sample in NMR buffer supplemented with 150 mM KCl. The sample pH was varied by gradual addition of a dilute HCl solution (pH ~ 3) and measured before and after each HSQC experiment. Changes in the amide chemical shifts of E535 and surrounding residues (Y536, L533 and A534) as a function of decreasing pH were monitored using standard ¹H-¹⁵N HSQC experiments (900 MHz). Changes in proton (δ_H) or nitrogen (δ_N) chemical shifts as a function of pH were fitted to a pseudo two-state equilibrium model using the following equation:

$$\delta_{\text{obs}} = (\delta_A + \delta_B 10^{n(\text{pH} - \text{pK}_a)}) / (1 + 10^{n(\text{pH} - \text{pK}_a)}),$$

where δ_{obs} is the observed ^1H (δ_{H}) or ^{15}N (δ_{N}) chemical shift, δ_A and δ_B are the fitted low and high pH chemical shifts respectively, pK_a is the apparent pK_a for the two-state equilibrium, and n is a cooperativity factor (fit values $1.5 < n < 2$). The pH titration curves and corresponding best fits and apparent pK_a values are shown in Figure S1.

NMR spin relaxation experiments were performed at 500 MHz using a ~ 0.2 mM sample and standard longitudinal (R_1) and transverse (R_2) pulse sequences. Data were acquired with 32 or 64 scans and time delays of 20, 60, 120, 200, 360, 600, 1000 (2X), 1400 ms for R_1 and 6 (2X), 10, 18, 26, 42, 58, 82 (2X), 106 ms for R_2 . Relaxation rates were determined by fitting the peak intensity as a function of time to a single exponential decay using Mathematica (Wolfram Research). Rotating frame ($R_{1\rho}$) relaxation dispersion (RD) profiles were measured at a single field (18.8 T, 800 MHz) using previously published ^{15}N experiments, which employ heteronuclear polarization transfers for selective excitation and a 1D acquisition scheme for probing individual nuclei.^{97,98} Relaxation data were collected with 128 or 256 scans and two time delays of 0 and 24 ms for E535 and Y536 in Kaiso-KBS and Kaiso-KBS(5TCT) and 0 and 20 ms for A536 in Kaiso(Y536A)-KBS. On-resonance RD data were recorded by varying the spin-lock power (ω_{SL}): 100 (2X), 150, 200, 250, 300, 400, 500, 600, 700, 800, 1000, 1200, 1400, 1600, 1800, 2000 Hz. Off-resonance RD data were recorded by varying the offset from the spin-lock carrier (Ω) at three to four different spinlock powers ranging from 200 to 1600 Hz. The offsets were $\pm 25, 50, 75, 100, 150, 200, 250, 300, 450, 600$ Hz for $\omega_{\text{SL}} = 200$ Hz; $\pm 50, 100, 200, 300, 400, 500, 600, 800, 1000, 1200$ Hz for $\omega_{\text{SL}} = 400$ Hz; $\pm 100, 200, 300, 400, 600, 800, 1000, 1400, 1800, 2400$ Hz for $\omega_{\text{SL}} = 800$ Hz and $\pm 100, 200, 400, 600, 800, 1200, 1600, 2200, 3000, 4200$ Hz for $\omega_{\text{SL}} = 1400$ Hz or 1600 Hz. $R_{1\rho}$ values were fitted to the Laguerre approximation model,⁹⁹ assuming a two-state equilibrium between a ground state (A) and an excited state (B):

$$R_{1\rho} = R_1 \cos^2 \theta + R_2 \sin^2 \theta + \sin^2 \theta p_A p_B \Delta \omega^2 k_{\text{ex}} / [\omega_A^2 \omega_B^2 / \omega_{\text{eff}}^2 + k_{\text{ex}}^2 - \sin^2 \theta p_A p_B \Delta \omega^2 B (1 + 2k_{\text{ex}}^2 (p_A \omega_A^2 + p_B \omega_B^2) / (\omega_A^2 \omega_B^2 + \omega_{\text{eff}}^2 k_{\text{ex}}^2))],$$

where $\omega_{\text{eff}}^2 = \Omega^2 + \omega_{\text{SL}}^2$, $\omega_A^2 = (\Omega_A - \omega_{\text{rf}})^2 + \omega_{\text{SL}}^2$, $\omega_B^2 = (\Omega_B - \omega_{\text{rf}})^2 + \omega_{\text{SL}}^2$, $\omega_B = \Omega_B - \Omega_A$, $\theta = \tan^{-1}(\omega_{\text{SL}} / \Omega)$, $\Omega = \Omega_{\text{ave}} - \omega_{\text{rf}}$. R_1 and R_2 are the longitudinal and transverse relaxation rates, respectively, Ω_A and Ω_B are the resonance offsets from the spin-lock carrier for the respective states, ω_{rf} and ω_{SL} are reference frequency and the strength of the spin-lock carrier, $\Omega_{\text{ave}} = p_A \Omega_A + p_B \Omega_B$, k_{ex} is the exchange rate, all in units of s^{-1} . Mathematica (Wolfram Research) was used for analysis and visualization of pH titration, R_1 and R_2 relaxation, and $R_{1\rho}$ RD data.

Results

Solution conformation and dynamics of E535 in Kaiso bound to specific DNA

In previous studies it was shown that Kaiso E535 is critical for binding to methylated DNA but is dispensable for binding to the KBS motif.⁸⁸ This observation appears to be inconsistent with subsequent high-resolution X-ray structures of Kaiso-DNA complexes, which showed that E535 adopts a similar conformation and is directly involved in recognition of TpG (in KBS) and mCpG sites, albeit with differences in the state of protonation of the E535 carboxyl.⁸⁷ The glutamate side chain is deprotonated in complexes with methylated DNA and forms a direct O---H-N hydrogen bond to each mC N4 amino group of the mCpG site, while it appears to be protonated in the complex with specific DNA, allowing it to form one O---H-O hydrogen bond to the T O4 carbonyl group and another O---H-N hydrogen bond to the C N4 amino group of the TpG site (Figure 1A). However, NMR spectra show a much larger difference in the amide ¹⁵N chemical shifts of E535 (~ 5 ppm) and the neighboring Y536 (~ 4 ppm) between the Kaiso-KBS and Kaiso-MeDNA complexes⁸⁷ than would be expected from differences in carboxyl protonation alone (~1 ppm),¹⁰⁰ suggesting possible conformational differences in solution.

We used NMR spectroscopy to probe whether the solution conformation of E535 in the Kaiso-KBS complex at neutral pH is different from that observed in the crystal structure (4F6M), where the close approach (2.7 Å) of the thymine O4 (T8 in the KBS oligonucleotide used in the present work, Table S1) and E535 Oε2 atoms implies protonation of the carboxyl and formation of a hydrogen bond (Figure 1A, green structure).⁸⁷ To elucidate the function of E535, we monitored the pH dependence of the backbone amide chemical shifts using ¹H-¹⁵N HSQC spectra of the Kaiso-KBS complex. As the pH was decreased, the cross peaks for E535 and Y536 (and to a lesser extent, L533 and A534) were observed to shift and broaden beyond detection, then reappear ~ 5.6 ppm upfield for E535 and ~ 4.0 ppm downfield for Y536 in the ¹⁵N dimension (Figure S1A). Fitting of the pH titration curves to a two-state model yielded an apparent pK_a of 5.5 ± 0.1 for all of these resonances (Figure 1B, S1B), suggesting that all reflect the same protonation event. The pH-dependent behavior at residues 533–536 was abolished when E535 was mutated to glutamine (Figure S1C). Thus, we reasoned that the pH-dependent chemical shifts reflect protonation of E535 at low pH and that E535 remains largely deprotonated in Kaiso-KBS at neutral pH. However, the large chemical shift changes at E535 and, especially, at Y536 cannot be explained by glutamate protonation alone¹⁰⁰ and imply a concomitant pH-dependent conformational change. The observed pK_a is also significantly higher than the intrinsic pK_a for a glutamate carboxyl group (~ 4.2),¹⁰⁰ consistent with the location of E535 in the protein-DNA interface.

The ¹⁵N chemical shifts of E535 and Y536 in Kaiso-KBS at low pH (Figure 1C, purple), where E535 is protonated, closely match those of the corresponding residues in spectra of Kaiso bound to different CpG-methylated DNA duplexes (MeDNAs) at neutral pH, where E535 is deprotonated (Figure 1C, salmon, blue, yellow cross peaks within the dotted ellipses). The similarity of the ¹⁵N chemical shifts provides strong evidence that E535 and Y536 are in similar structural environments in the low-pH Kaiso-KBS complex and the

neutral-pH Kaiso-mCpG complexes. Additional insights are obtained from spectra of the Kaiso(E535Q)-KBS complex, where the negatively charged E535 carboxyl is replaced with a neutral carboxamide group that can potentially form a direct hydrogen bond to the T8 O4 in the KBS oligonucleotide (Table S1).^{53,87} The backbone amide cross peaks of Q535 and Y536 in Kaiso(E535Q)-KBS at neutral pH (Figure 1C, green) are shifted in the ¹⁵N dimension by ~ 4.0 and ~ 2.4 ppm, respectively, towards their positions in spectra of Kaiso-MeDNA complexes or Kaiso-KBS at low pH. This suggests that the conformation of Q535 resembles that of E535 in Kaiso-KBS at low pH or in Kaiso-MeDNA. Conversely, the spectrum of the Kaiso-KBS complex at neutral pH features drastically different chemical shifts for E535 and Y536 (Figure 1C, grey), indicating a different conformation at that site.

The pH titration of Kaiso-KBS showed pronounced cross peak broadening for E535 and Y536 (Figure S1), which could result from conformational dynamics on the micro-to-millisecond (μ s-ms) timescale. We performed ¹⁵N spin relaxation experiments to probe the Kaiso backbone dynamics on timescales from picoseconds (ps) to milliseconds (ms). We observed significantly increased transverse relaxation rates (R_2) for E535 and Y536 relative to other residues in Kaiso-KBS at pH 7.0 (Figure S2), indicating μ s-ms exchange dynamics at those sites. By contrast, these residues do not have elevated R_2 values in the Kaiso(E535Q)-KBS and Kaiso-MeCG2 complexes (Figure S2), consistent with lack of chemical exchange. We also computed the R_2/R_1 and $2R_2-R_1$ values, which scale roughly with overall molecular tumbling and fast picosecond-to-nanosecond (ps-ns) internal motions, respectively¹⁰¹ (Figure S2). Both the R_2/R_1 and $2R_2-R_1$ values were generally lower in Kaiso-KBS relative to the E535Q mutant complex (except for sites of chemical exchange), particularly in Kaiso ZF3 and C-terminal tail regions, which suggests increased fast dynamics in the wild-type complex.

To characterize the chemical exchange process in Kaiso-KBS, we employed rotating-frame ($R_{1\rho}$) relaxation dispersion (RD), which can probe μ s-ms exchange between an observed major state and a sparsely populated, invisible state.^{102,103} Specifically, we used a 1D selective ¹⁵N $R_{1\rho}$ off-resonance RD experiment, which generates extensive RD profiles as a function of spin-lock strength (ω_{SL}) and resonance offset (Ω);⁹⁷ these profiles can be fit to chemical exchange models that yield quantitative information about the exchange rate and the population and ¹⁵N chemical shift of the minor state. RD profiles recorded for E535 and Y536 are shown in Figure 1D; the data were best fit to a global two-state model (Table S3). We found that, at neutral pH, E535 and Y536 each interconvert rapidly ($k_{ex} \sim 8400 \pm 600 \text{ s}^{-1}$) between a major deprotonated state A and a minor protonated state B (population $p_B \sim 5.5 \pm 0.5 \%$). The ¹⁵N chemical shifts of the minor state are 5.5 ± 0.5 ppm upfield for E535 and 4.3 ± 0.4 ppm downfield for Y536 (arrows in Figure 1C). These values agree with the corresponding chemical shifts observed directly at low pH in Kaiso-KBS Figure 1C, purple) and at neutral pH in Kaiso-MeDNA (Figure 1C, dotted ellipses), implying that the chemical exchange process reflects the protonation equilibrium of the glutamate carboxyl group and its associated conformational changes. The observed kinetic rate constant for deprotonation (k_B) can be derived from k_{ex} and the population of the major state p_A ($k_B = k_{ex}p_A$). The k_B , $8000 \pm 600 \text{ s}^{-1}$, is substantially slower than proton dissociation rates ($\sim 10^5 - 10^6 \text{ s}^{-1}$) measured by NMR relaxation for exterior and interior glutamates in small proteins,¹⁰⁴ but is comparable to the proton dissociation rate for an interior glutamate with associated local

loop dynamics ($\sim 1-9 \times 10^3$),¹⁰⁵ further supporting the coupling of E535 protonation to a conformational transition.

The exchange process was abrogated when E535 in Kaiso-KBS was mutated to glutamine (E535Q) or alanine (E535A) or when wild-type Kaiso was bound to methylated DNA (Figure S3A). However, the chemical exchange persisted in Kaiso Y536A bound to KBS, indicating that the conformational exchange is driven by E535 and not by Y536 (Figure S3B). Taken together, the solution NMR data strongly suggest that E535 in the Kaiso-KBS complex adopts a predominantly deprotonated state at physiological pH in solution, in contrast to the protonated state inferred from the 4F6M crystal structure.⁸⁷

Distinct E535 conformations in structures of Kaiso bound to KBS and methylated DNA

The conformation of E535 in Kaiso complexes with KBS DNA was determined by X-ray crystallography. To visualize the deprotonated state that E535 adopts at neutral pH, the structure of Kaiso-KBS was solved at pH 7.2 (1.8 Å resolution) (Figure 2A, Table S2). The structure at pH 7.2 revealed a dramatic shift in the position of E535 compared to the previously reported structure (4F6M, Figure 1A).⁸⁷ The latter structure was determined using crystals obtained from unbuffered 0.2M sodium nitrate, 20% PEG3350; the pH of the Kaiso-KBS crystallization medium is unknown, but the structure is that of the low-pH form. In the pH 7.2 structure, the E535 side chain is flipped away from the TpG site, assuming a conformation (“out” state) with no direct contacts to the DNA. Instead, E535 forms water-mediated hydrogen bonds with the T8-A29 base pair of the TpG site and the 5' neighboring C7 N4 amino group (Figure 2A). Lowering the pH of the Kaiso-KBS crystallization buffer to 6.5 yielded a 2.6 Å crystal structure (Figure 2B, Table S2), which shows E535 in the same conformational state as previously observed.⁸⁷ In this state, henceforth termed the “in” state, the E535 carboxyl appears to be protonated and makes direct hydrogen bonds with T8 O4 and C28 N4 in the TpG site (Figure 1A). Interestingly, we consistently observed the protonated “in” state in two independent crystal structures of Kaiso-KBS at pH 6.5, even though it is not predicted to be predominant in solution at this pH, based on the apparent pK_a of E535 and the chemical exchange dynamics obtained from NMR studies above.

In the Kaiso-KBS structure at pH 7.2, the side chain of E535 in the “out” conformation makes a direct O---HN hydrogen bond (Oe2---N ζ 2.7 Å) to the side chain of K539, as well as water-mediated hydrogen bonds with C7 N4 and A29 N6 and N7 (Figure 2A). The conformation of K539 differs in the low-pH Kaiso-KBS structure, where the lysine amino group points towards the backbone phosphate groups of C28 and A29 (Figure 2B). To determine whether the interaction between E535 and K539 stabilizes the “out” over the “in” state of E535, we engineered a Kaiso K539A mutant and solved its structure (at 1.7 Å resolution) bound to KBS at pH 7.2 (Figure 2C, Table S2). With the K539A mutation, E535 now preferentially adopts the protonated “in” state over the deprotonated “out” state. We also monitored the behavior of Kaiso K539A bound to specific and methylated DNA in solution using NMR. The N-H cross peak of E535 is broadened beyond detection in spectra of the Kaiso(K539A)-KBS complex at neutral pH but can be observed for the protonated state at pH 5.4 (Figure 2D, red); in the wild-type complex, the cross peak of protonated E535 is observable only at pH 4.8 (Figure 1B, Figure 2D, purple cross peak). The cross-

peaks for adjacent residues (Y536, A534, and L533) in Kaiso(K539A)-KBS at pH 7 were broadened but observable and found at chemical shifts corresponding to the protonated E535 “in” state (Figure S4A), indicating that this is the major glutamate conformation. These results are consistent with the crystal structures and imply an increase in the apparent pK_a of E535 caused by the lysine mutation. By contrast, no line broadening and only a small chemical shift change was observed for E535 cross peak in spectra of the WT (yellow) and K539A (blue) complexes with MeDNA (Figure 2D). These findings support the hypothesis that the hydrogen bond between E535 and K539 helps stabilize the “out” conformation of deprotonated E535.

We further hypothesized that mutation of E535 to glutamine would favor the “in” state since Q535 does not require protonation to form a direct hydrogen bond to T8 O4 at the TpG site in KBS. This is consistent with the similarity of the positions of the backbone amide cross peaks of Q535 in spectra of Kaiso(E535Q)-KBS (Figure 1C, green) and E535 Kaiso-KBS at low pH (Figure 1C, purple) or in Kaiso-MeDNA (Figure 1C, teal, yellow, salmon), where the “in” state is directly observed (Figure 1A, Figure 2B). The crystal structure of Kaiso(E535Q)-KBS at pH 7.2 (2.3 Å) (Table S2) confirms that Q535 adopts the “in” state, where the carboxamide group forms direct hydrogen bonds to T8 O4 and C28 N4 in the TpG site (Figure 2E). These interactions are analogous to the hydrogen bonds between E535 and the TpG site at low pH (Figure 2B).

Mutation of E535 to alanine was previously shown to preserve Kaiso binding to KBS but abolish binding to methylated DNA.⁸⁸ This differential effect on binding can now be explained by the different conformations of E535 in Kaiso bound to KBS (E535 “out”) and MeDNA (E535 “in”) at physiological pH. To investigate how the E535A mutation affects Kaiso-KBS interactions and explain the small effect of E535A on KBS binding, we determined a crystal structure of the Kaiso(E535A)-KBS complex at pH 7.2 (1.9 Å) (Figure 2F, Table S2). Despite the absence of the E535 side chain, the hydration network at the protein-DNA interface is largely preserved. Water molecules that bridge interactions between E535 and the T8-A29 base pair and the 5' flanking C7 in the wild-type complex remain bound at similar positions in the mutant complex and form part of an extensive hydrogen bonded water network within the protein-DNA interface (Figure 2F). In the absence of the glutamate side chain, K539 reverts to the conformation adopted at low pH (Figure 2B).

E535 conformation modulates Kaiso-DNA binding affinity and sequence specificity

To probe the role of E535 in discriminating between sequence-specific (KBS) and methylated DNA sites, we interrogated the DNA-binding properties of Kaiso wild-type and E535 mutants. We employed an *in vitro* biolayer interferometry (BLI) assay that measures the kinetics of protein binding to biotinylated DNA immobilized on a streptavidin-coated sensor (representative data are shown in Figure S5). The kinetic data were fitted to a two-state binding model to obtain the association (k_{on}) and dissociation (k_{off}) rate constants, which were used to calculate the dissociation constant for protein-DNA binding (K_d) (Table 1). The BLI data revealed that Kaiso binds tightly, with low nanomolar affinity ($K_d \sim 2.8 \pm 0.4$ nM), to the KBS motif under conditions (pH 7.0 and 150 mM KCl) that mimic

physiological pH and salt. This value is ~15-fold higher than the K_d reported previously using a gel-shift assay under low-salt conditions (180 ± 30 pM).⁵⁴ Mutation of E535 to glutamine (E535Q) led to a 5-fold enhancement in KBS binding affinity ($K_d \sim 0.53 \pm 0.07$ nM) relative to the wild-type protein (Table 1), while mutation to alanine (E535A) had a negligible effect on the affinity for KBS ($K_d \sim 4.0 \pm 0.7$ nM) (Table 1). Changes in binding affinity were primarily due to changes in the dissociation rate constants (k_{off}). These data are consistent with the structural studies, which show that E535Q gains two direct hydrogen bonds between the glutamine and the core TpG site (Figure 2E), while E535A does not significantly perturb the protein-DNA interface or hydration network (Figure 2F).

Wild-type Kaiso binds ~ 14-fold more weakly to KBS than to MeKBS DNA (K_d 0.20 ± 0.04 nM) (Table 1), as would be expected given the lack of direct E535-TpG interactions in the predominant E535 “out” conformation at pH 7.⁵³ Remarkably, the specificity is reversed in favor of KBS binding by both the E535Q (14-fold) and E535A (9-fold) mutations. The glutamine mutant enhances binding to the specific KBS DNA, consistent with the gain of direct Q535-TpG hydrogen bonds, while lowering affinity for methylated DNA (> 30-fold) where one of these hydrogen bonds is broken.⁵³ The E535A mutant, which lacks these hydrogen bonds with either DNA substrate, inhibits binding to methylated DNA even more dramatically (>100-fold) but has a negligible effect on binding to KBS. These results underscore the importance of the E535 carboxyl group for recognition of and high affinity binding to mCpG sites and suggest a potential approach to modulate Kaiso binding to sequence-specific versus methylated DNA target sites *in vivo*.

Even though E535 in the “out” conformation is dispensable for KBS binding, indirect water-mediated contacts with the DNA suggest that it may affect sequence specificity and readout. Early studies of Kaiso-DNA recognition using a SELEX approach identified a consensus Kaiso binding site (5'-TCCTGCNA) with a highly conserved CTGCNA core and showed that the protein has a preference for cytosine immediately 5' to the core TpG site.⁴⁵ To investigate the role of E535 in recognition of the 5' flanking sequence, we measured the affinity of wild-type Kaiso and E535 mutants for KBS variants where the C7-G30 base pair in KBS was substituted with a G-C (KBS(5TCG)), A-T (KBS(5TCA)) or T-A (KBS(5TCT)) base pair (Figure 3A). The binding kinetics and affinities are reported in Table 1. The wild-type protein showed significant loss in binding affinity with all KBS variants, ~ 20-fold for KBS(5TCG) and ~ 80-fold for KBS(5TCA) and KBS(5TCT) (Figure 3B). The E535A mutant showed a similar trend, but the difference in K_d was roughly 3 times smaller than with the wild-type protein, due to a slightly reduced affinity for KBS but increased affinity for KBS variants (Figure 3B). The E535Q mutant binds with higher affinity to the variant KBS sequences than does wild-type Kaiso, resulting in an even weaker ability to discern the 5' flanking base pair. Notably, Kaiso E535Q exhibited virtually the same binding affinity for KBS and KBS(5TCG). Thus, while E535A and E535Q mutations have differing effects on the binding of Kaiso to the KBS site and its variants, both partially impair the ability of Kaiso to discriminate against non-cognate 5' sequences.

We also examined the binding of Kaiso to 5' base pair variants in the context of CpG-methylated DNA (Table 1). Based on *in vitro* biochemical and *in vivo* genomic studies of Kaiso binding to methylated DNA motifs, the protein can recognize mCpG sites embedded

in variable sequence contexts.^{45,85,86} This is confirmed by the present experiments, where modification of the 5' or 3' base pairs flanking the mCpG site (C-G and A-T, respectively, in MeKBS) in the MeCG2, MeCG2B, and MeEcad sequences results in only a 2- to 6-fold decrease in binding affinity (Figure 3, Table 1). The small change in affinity associated with changes in the mCpG flanking sequences is in contrast to the dramatic, as much as two orders of magnitude, decrease in affinity caused by substitution of the 5' base pair preceding the KBS sequence.

Since K539 modulates the conformational equilibrium at E535, we also probed the effect of a K539A mutation on Kaiso binding to specific and methylated DNA (Table 1). Mutation of the lysine led to a 3-fold decrease in affinity for wild-type KBS but much greater effects for some of the KBS variants. Even though the K539A mutation tilts the E535 equilibrium towards the protonated “in” state that makes two direct hydrogen bonds to the TpG site, we do not observe an increase in binding affinity. The K539A mutant has a similarly small effect (2 to 3-fold) on binding to most methylated DNA sequences studied, with the exception of MeCG2B for which a ~ 15-fold decrease in affinity was observed. Next, we examine in detail the K539 conformation and DNA interactions with specific and methylated motifs to gain insight into these effects.

Dependence of E535 and K539 conformation on the 5' flanking base pair

The conformation of E535 in Kaiso complexes with KBS variants was examined by solution NMR to probe the sources of Kaiso selectivity for the 5' flanking sequence. The amide cross peak of E535 could not be observed for any of the complexes with KBS variants, and, unlike complexes with methylated DNA (Figure 1C), severe line broadening was observed for the neighboring Y536 and A534 amides (Figure 3C). This hinted at a possible shift in the apparent pK_a for E535 (concomitant with increase in the protonated population, p_B) or, alternatively, an increase in the chemical exchange rate (k_{ex}) or the chemical shift difference (ω) between major (deprotonated) and minor (protonated) states. To test this, we increased the pH and measured chemical exchange using $R_{1\rho}$ relaxation dispersion. Under higher-pH conditions, the line broadening at Y536 and A534 was largely alleviated in the Kaiso-KBS(5TCT) complex (pH 8.2) but remained substantial in Kaiso-KBS(5TCG) (pH 8.5); the amide cross peak of E535 could not be observed in either of the spectra (Figure 3C). To obtain information on the E535 proton exchange process, we collected ^{15}N $R_{1\rho}$ RD data for Y536 in Kaiso-KBS(5TCT) and extracted the chemical exchange parameters by assuming correlated dynamics between E535 and Y536, as observed in Kaiso-KBS (Figure 3D, Table S3). The k_{ex} values determined for the Kaiso-KBS(5TCT) (pH 8.2) and Kaiso-KBS (pH 7.0) complexes are the same within experimental error. The population of the protonated “in” state in Kaiso-KBS(5TCT) at pH 8.2 (~ 3.5%) is larger than the ~ 0.2% calculated for Kaiso-KBS at this pH based on the measured pK_a of E535 (Figure 1B). This suggests that E535 experiences a shift to higher apparent pK_a when Kaiso is bound to KBS(5TCT). Based on greater line broadening with KBS(5TCG) (Figure 3C), we anticipate the apparent pK_a to be even more shifted than for KBS(5TCT). These results show that substitution of the 5' flanking C-G base pair perturbs the equilibrium between the protonated “in” and deprotonated “out” states of E535; the lowest apparent pK_a is with a 5' C and increases with

a 5' T or G nucleobase. This trend suggests that electrostatics at the interface of Kaiso with the 5' flanking sequence in KBS may significantly impact sequence specificity.

We also examined the effect of a K539A mutation on the conformation of E535 in the presence of KBS DNA variants. Kaiso(K539A)-KBS favors the E535 “in” state by eliminating the direct K539-E535 hydrogen bond (Figure 2C). Comparison of NMR spectra of Kaiso K539A bound to KBS, KBS(5TCT), and KBS(5TCG) at pH 7 shows that the amide cross-peaks of Y536, A534 and L533 are substantially shifted towards their expected chemical shifts when E535 occupies the protonated “in” state (Figure S4); the E535 cross peak could not be observed. The line broadening is similar across all complexes, except for Y536 in the spectrum of the Kaiso(K539A)-KBS complex, which might be explained by the larger chemical shift difference between “in” and “out” states (Figure S4A). These data show that E535 switches to the protonated “in” state in Kaiso K539A complexes and suggests a narrower range of apparent E535 pK_as between KBS variants than is observed in wild-type complexes.

To further elucidate how the different conformation of E535 in specific and methylated complexes might enable different sensitivity to 5' flanking sequences, we inspected available crystal structures of Kaiso-DNA complexes. We observed that K539, which interacts with E535 in the “out” state in Kaiso-KBS, is free to adopt various alternate conformations when E535 is in the “in” state at low pH or in complexes with methylated DNA. In Kaiso-MeKBS, the K539 side chain points towards the backbone of mC28 and G29 (Figure 4A) and makes water-mediated hydrogen bonds with their phosphate groups.

A similar conformation is observed in Kaiso(E535A)-KBS (Figure 2F) and, presumably, in Kaiso-KBS at low pH (Figure 2B), where the water is not observed but the K539 conformation is similar. In Kaiso-MeCG2, which contains a 5' flanking T-A base pair, the K539 side chain protrudes deep into the major groove and directly hydrogen bonds with the thymine base (T7 O4) (Figure 4B). In Kaiso-MeEcad, which features a different 5' flanking sequence (Figure 3A), K539 extends beyond the 5' flanking base pair to hydrogen bond with the backbone phosphate of A6 (Figure 4C). We also sought to determine structures for Kaiso bound to MeCG2B containing a 5' A-T base pair as well as 5' KBS variants. However, we only successfully crystallized and solved the structure of Kaiso bound to KBS(5TCG) at pH 5.0 (2.1 Å). There, E535 adopts the “in” state and K539 hydrogen bonds with G7 O6 in the major groove, analogous to its interaction with T7 O4 in Kaiso-MeCG2 (Figure 4D). Therefore, K539, when not constrained by E535, is able to establish specific (base) or non-specific (backbone) interactions with different 5' flanking sequences. Such diversity in K539-DNA interactions could contribute to the lower sequence specificity of Kaiso for methylated DNA. Moreover, hydrogen bonding between K539 and a 5' G or T in KBS variants could compete with hydrogen bonding between K539 and E535 and shift the equilibrium towards the “in” state.

Molecular basis for different Kaiso readout of the 5' flanking sequence in KBS and methylated DNA

Examination of Kaiso-KBS crystal structures shows that the 5' flanking C7-G30 base pair is sandwiched between the side chains of two protein tyrosine residues, Y536 and Y562

(Figure 5A). The aromatic ring of Y536 forms hydrophobic contacts with the C6 sugar and the aromatic C7 base but there is no appreciable π - π stacking. The aromatic ring of Y562 forms hydrophobic contacts with the A29 sugar and stacks in an edge-to-face (T-shaped) manner with the G30 base. The hydroxyl groups of both Y536 and Y562 also form direct hydrogen bonds to the backbone phosphate groups of C7 and G30, respectively. There is a pronounced kink in the DNA at the site of the tyrosine interactions, which seems to be generated by a positive roll and a reduced twist and rise at the C7pT8 base pair step. This also correlates with widening of the DNA minor groove by ~ 3 Å at that site which persists into the core TGCC site (Figure S6). However, we do not observe simultaneous compression of the major groove, whose width is largely constrained by insertion of the Kaiso ZF domains (Figure S6). This deformation gives rise to an overall bending of the DNA helix that appears to be stabilized through interactions with Y536 and Y562 and, possibly, additional base-pair specific (Q563 in ZF3) and backbone contacts (H543 in ZF2) (Figure 5C, Figure S7).

Substitution of the 5' C-G base pair in MeKBS with a 5' T-A base pair in MeCG2 inserts a methyl group at the interface with Y536.⁵³ This causes the DNA backbone and the thymine ring to be shifted away from the tyrosine ring, disrupting hydrophobic contacts to the C6 sugar and a direct hydrogen bond between the tyrosine hydroxyl and T7 phosphate (Figure 5A). Moreover, we observe two alternate DNA conformations in the 5' flanking sequence in Kaiso-MeCG2 – one that adopts a similar helical bend as in Kaiso-KBS (state A) and one that is less bent (state B) – which make distinct interactions with two alternate conformations (“down” and “up”) of Q563 in ZF3 (Figure 5C, Figure S7). State A is compatible only with the Q563 “down” conformation, which is similar to Kaiso-KBS but can no longer form a direct hydrogen bond to G31 O6 due to the helix displacement; state B is coupled to the Q563 “up” conformation, which makes hydrophobic contacts with the T4 base. At the same time, the major and minor groove widths are largely preserved. By analogy, we expect that a clash between Y562 and the thymine methyl group in a 5' A-T base pair would similarly cause DNA helical distortion or structural heterogeneity and result in a loss of hydrogen bonding and π - π stacking between Y562 and DNA. Structural modeling of Kaiso complexes with KBS substituted with a 5' T-A or 5' A-T base pair confirms that maintaining the same DNA conformation as observed in the wild-type Kaiso-KBS complex would lead to a steric clash between the thymine methyl group and the aromatic ring of Y536 or Y562, respectively (Figure 5B). Although we failed to obtain crystals of Kaiso-KBS(5TCT), the HSQC spectrum shows that cross-peaks of many residues near the interface with DNA are shifted towards their positions in the spectrum of Kaiso-MeCG2, suggesting that the complexes may have similar conformations (Figure S8A). Therefore, similar DNA deformations, coupled with loss of hydrogen bonding and hydrophobic contacts, as in Kaiso-MeCG2, may be required to alleviate the tyrosine-methyl steric clash in Kaiso complexes with KBS(5TCT) and KBS(5TCA), which could explain the reduced binding affinities relative to the consensus site (Table 1).

To test this hypothesis, we substituted the 5' T immediately preceding the TpG site in KBS(5TCT) and KBS(5TCA) with a uridine, which lacks the methyl group but otherwise forms identical Watson-Crick hydrogen bonds with the complementary adenine. First, we measured the binding kinetics of Kaiso for the uridine-containing KBS(5TCU/A) and

KBS(5TCA/U) constructs using the BLI assay. We observed a substantial increase in Kaiso binding affinity for the 5' U-A (~17-fold) and A-U (~10-fold) base pair as compared to their thymine counterparts, 5TCT and 5TCA (Table 1). Remarkably, the K_{ds} for KBS(5TCU/A) and KBS(5TCA/U) were now only 4-fold and 8-fold weaker, respectively, than the K_d for KBS (Figure 5D) as opposed to roughly 80-fold weaker with the thymine base at this site (Figure 3B). We further interrogated the binding of Kaiso E535 mutants to the uridine-substituted DNAs and found that their binding affinities were also enhanced relative to the thymine analogs, but to different extents (Table 1). These findings show that removal of the thymine methyl group in KBS(5TCT) and KBS(5TCA) increases binding affinity and reduces Kaiso selectivity towards the 5' flanking base pair; mutation of E535 to alanine or glutamine decreases selectivity even further (Table 1). Wild-type Kaiso binds more tightly to the KBS(5TCU/A) and KBS(5TCA/U) over the KBS(5TCG) motif and the order of binding affinities from strongest to weakest is 5' C > U/A > A/U > G. The E535A and E535Q mutants display different sequence specificities than the wild-type protein (5' C > U/A > G > A/U for E535A; 5' G ~ C > U/A ~ A/U for E535Q), suggesting that the wild-type glutamate is involved in discrimination of the 5' base-pair.

To probe how the T-to-U substitution affects the protein-DNA conformation, we determined the crystal structure of wild-type Kaiso bound to KBS(5TCU/A) at pH 7.2 (1.9 Å). The DNA and protein conformation, both globally and in the vicinity of the U-A base pair (Figure 5B), matched closely that of the pH 7.2 Kaiso-KBS complex, with E535 in the “out” conformation. The van der Waals and hydrogen bonding contacts of Y536 and Y562 with DNA remained largely unperturbed (Figure 5B). Importantly, the water network at the interface with the DNA 5' flanking sequence was nearly identical, including conserved water molecules that form bridging hydrogen bonds between E535 and the 5' C7-G30 (KBS) or 5' U7-A30 (KBS(5TCU/A)) base pair. ^1H , ^{15}N -HSQC spectra of Kaiso bound to KBS(5TCU/A) (Figure S8B, magenta) showed that the backbone amide chemical shifts matched very closely those of the Kaiso-KBS complex (Figure S8B, black), indicating a protein conformation more similar to the KBS complex than to KBS(5TCT) (Figure S8B, green). However, we also detected line broadening for Y536 as we did with Kaiso-KBS(5TCT), suggesting that the apparent pK_a of E535 may be shifted in favor of the “in” state (the E535 cross peak in the KBS(5TCT) and KBS(5TCU/A) complexes could not be resolved due to broadening or overlap) (Figure S8B). This is not surprising since, apart from the methyl group, the T and U nucleobases are nearly identical and have similar major groove hydrogen bonding capabilities and electrostatics. Together, our structural and binding studies show that the T-to-U substitution restores binding by alleviating steric clash and accommodating a similar DNA conformation and Y536-DNA contacts as in the Kaiso-KBS complex. The aromatic rings of Y536 and Y562 thus play an important role in discriminating between C-G and T-A/A-T base pairs adjacent to the core TpG site in KBS DNA.

Finally, we investigated whether the Kaiso Y536 and Y562 side chains are important for recognition of sequence-specific versus methylated DNA motifs. Using the BLI assay, we observed that the Y536A mutation impaired binding to KBS (~100-fold) and led to large losses in the binding affinity for KBS variants (Table 1). The Y562A mutation resulted in an even greater loss in binding affinity for KBS and no distinguishable sequence specificity

(Table 1). The greater effect of Y562A could be due to disruption of a hydrophobic cluster formed by residues Y562, Y584, Y550, M565, and P577, where all three tyrosines form direct hydrogen bonds with the DNA backbone.⁸⁷ We also wondered whether these mutations would have a similar impact on binding to methylated DNA. Surprisingly, the Y536A mutation had little to no effect (< 3.5-fold) on binding to methylated DNA containing different 5' flanking base pairs (Table 1). The Y562A mutation appreciably weakened binding to methylated DNA (~ 15 to 25-fold), although to a much smaller extent than for KBS (~ 240-fold) (Table 1). These data show that the tyrosine mutations, especially Y536A, selectively destabilize complexes with specific KBS DNA sequences, which could result from disruption of Y536-DNA contacts and possibly an altered DNA conformation that could affect other protein-DNA interactions. This is clearly not the case for methylated DNA, where the tyrosine mutations have much smaller effect on binding. Thus, our findings suggest that Kaiso employs a different readout mechanism for sequence-specific versus methylated DNA sites.

Discussion

Until now, it has been unclear how Kaiso and other MBPs achieve different sequence specificity for methylated and consensus motifs and selectively target *in vivo* sites depending on their methylation status. Here, we show that Kaiso undergoes an E535 protonation-coupled conformational switch when binding to KBS versus CpG-methylated DNA. The two distinct conformations at E535 enable a different set of interactions that lead to different mechanisms of recognition: (i) direct readout of the 5' mCpG site by E535 that dominates the binding affinity in methylated DNA, and (ii) largely indirect readout of the TpG site in KBS DNA coupled with sequence-dependent differences in DNA flexibility or hydration.

A key discovery in the present work is the difference in the conformation of E535 in Kaiso complexes with specific versus methylated DNA, which directly impacts the DNA sequence preference and readout. Using solution NMR and X-ray crystallography, we show that E535 adopts a deprotonated ("out") conformation in the KBS complex at physiological pH. In stark contrast to the "in" conformation observed in the low-pH Kaiso-KBS and the Kaiso-MeDNA structure,^{53,87} E535 in the "out" conformation maintains only water-mediated hydrogen bonds with the DNA, and is stabilized by a hydrogen bond to the side chain of a nearby lysine (K539). In solution, E535 in the Kaiso-KBS complex interconverts between the major "out" and the minor "in" conformation on a microsecond timescale, but this equilibrium is tipped in favor of the protonated "in" state by the K539A mutation. The observation of two different E535 conformations in sequence-specific and methylated DNA complexes provides mechanistic insights into previous gel retardation studies⁸⁸ and our quantitative BLI data which show that mutation of E535 to alanine impairs binding to methylated DNA but not to the specific DNA motif. For methylated DNA, the loss in binding affinity results from loss of multiple direct (canonical and CH---O) hydrogen bonds and van der Waals contacts between the carboxyl group of E535 in the "in" conformation and the mC amino and methyl groups.⁵³ For KBS, however, the negligible effect of the E535A mutation on binding indicates that the water-mediated interactions between E535 in the "out" conformation and the specific site contribute little to the DNA binding affinity.

The conformational switch at E535 has a profound effect on the binding affinity and mechanism of DNA recognition. We find that the difference in E535 conformation can largely account for differences in affinity for KBS versus methylated DNA sites. Based on our BLI data, the affinity of Kaiso for KBS is more than an order of magnitude weaker than for a similar methylated DNA sequence (MeKBS), where the deprotonated E535 side chain adopts the “in” conformation and forms canonical hydrogen bonds with mC8 N4 and mC28 N4. The E535Q mutant, which effectively mimics the “in” conformation of E535 in crystal and in solution, forms hydrogen bonds to T8 O4 and C28 N4 and increases the affinity for the KBS motif to nearly the same levels as for wild-type Kaiso binding to MeKBS (Table 1).

Surprisingly, the E535Q mutation leads to a much smaller difference in the KBS binding free energy relative to the wild-type protein ($\Delta G \sim 1.0$ kcal/mol) or the E535A mutant ($\Delta G \sim 1.2$ kcal/mol) than would be expected based on the gain of two direct hydrogen bonds between the glutamine side chain and DNA (~ 2 kcal/mol¹⁰⁶). By contrast, the E535A mutation, which disrupts similar canonical hydrogen bonds and weak CH $\cdots$ O hydrogen bonds in the wild-type Kaiso-MeDNA complex, leads to a 3.1 kcal/mol change in ΔG . The difference in ΔG between complexes with Kaiso-specific and methylated DNA (~ 2 kcal/mol) could be attributed to the CH $\cdots$ O hydrogen bonds⁵³ and to more favorable electrostatic interactions between the charged glutamate and the mC base that could strengthen hydrogen bonds between them. In addition, differences in entropy (i.e. conformational dynamics) between the two complexes could also play a role. NMR relaxation data show pronounced μ s-ms time scale dynamics at E535 and neighboring residues in the wild-type Kaiso-KBS complex that are absent in the corresponding E535Q complex, together with generally lower $2R_2-R_1$ values in the wild-type complex that are indicative of enhanced ps-ns internal motions. It is possible that the lack of direct contacts between E535 and the TpG step in the wild-type Kaiso-KBS complex leads to increased protein and/or DNA mobility that impacts the binding affinity (i.e. reduces ΔG).

The different E535 conformation in Kaiso complexes with KBS and mCpG motifs also correlates with significant differences in sensitivity towards the neighboring DNA sequence. In contrast to complexes with methylated DNA, Kaiso is highly selective for the base pair 5' to the core TpG site in KBS (5' C-G). Thymine-containing base pairs (5' T-A and A-T) sterically clash with key Kaiso tyrosine residues (Y536 and Y562), which can largely be alleviated by substitution with a uridine base (see below). Even then, Kaiso binding to the non-cognate 5' U-A (4-fold), 5' A-U (8-fold), and 5' G-C (20-fold) base pairs is disfavored compared to the cognate 5' C-G base pair. This sequence preference could be rooted in unfavorable electrostatic interactions between the negatively charged glutamate side chain and a 5' nucleobase with a more electronegative major groove potential (5' C < T/U < A < G, from positive to negative).¹⁰⁷ In support of this, we find that the binding affinity after eliminating the tyrosine-thymine clash (K_d for 5' C < U < A < G) closely follows the electrostatic trend above. Moreover, mutation of E535 to glutamine diminishes the ability of Kaiso to discriminate against non-cognate 5' base pairs; affinity for the most electronegative guanine is enhanced with Kaiso E535Q to a point where the mutant protein does not discriminate between KBS and KBS(5TCG). Finally, NMR studies report a greater shift in the E535 conformational equilibrium away from the negatively charged “out” state with more electronegative nucleobases (5' C < T $\sim$ U < G), consistent with a shift in the apparent

pK_a of E535 due to changes in electrostatic potential in the major groove. Concomitantly, the E535 equilibrium could be modulated by the ability of K539 to extend into the DNA major groove and form direct hydrogen bonds with a 5' G or T/U base when E535 occupies the "in" conformation, as observed in crystal structures of Kaiso bound to MeCG2 or KBS(5TCG) at low pH (Figure 5). A greater propensity of K539 to interact with 5' G-C and 5' T/U-A base pairs would disfavor the E535 "out" conformation, in line with NMR data showing greater line broadening in KBS 5' variants. A K539A mutation appears to suppress the shift in apparent pK_a of E535 in Kaiso bound to KBS 5' variants and leads to larger effects on their K_d s, consistent with sequence-dependent lysine-DNA interactions.

Alternatively, the diminished affinity for binding to KBS sequences containing non-cognate 5' base pairs could also stem from intrinsic differences in DNA deformability. In particular, the C7pT8 step in KBS, where we observe large deviations from ideal DNA parameters in the form of increased roll and reduced twist (Figure S6), could be more prone to such deformations than are non-cognate steps. Indeed, CpT steps are generally considered to be more flexible than GpT, TpT or ApT steps.⁶² However, their flexibility is highly context dependent⁶³ and likely affected by replacement of T with U, where methyl-specific steric interactions with the sugar backbone and adjacent nucleotides¹⁰⁸ that restrict local motions are absent. Given that experimental measurements and theoretical predictions of DNA flexibility are limited, particularly for U-A base pairs incorporated in B-DNA, more detailed studies will be required to establish correlations between sequence selectivity of KBS sequences and inherent flexibility at the kink site.

Another important finding is that binding of Kaiso to specific, but not methylated sites, is highly sensitive to interactions between key aromatic residues, Y536 and Y562, and the DNA. The tyrosine side chains sandwich the 5' C-G base pair and stack in an edge-to-edge (Y536-C7) or edge-to-face (Y562-G29) manner with the nucleobases, while also forming hydrophobic and hydrogen bonding contacts with the sugar-phosphate backbone. These interactions appear to stabilize the kink at the C7pT8 step, resulting in bending of the DNA helix towards the major groove. Y562 is also part of a hydrophobic cluster containing two additional tyrosines (Y550 and Y584) that form direct hydrogen bonds with the DNA phosphate backbone. Mutation of either Y536 or Y562 to alanine greatly impairs binding to KBS. Moreover, binding of wild-type Kaiso to KBS variants (5TCT and 5TCA) containing 5' T-A or A-T flanking base pairs, which cause steric clash between the thymine methyl groups and tyrosine side chains, is impaired almost 100-fold (Table 1). Eliminating the steric clash with a T-to-U base substitution restores the binding affinity to within several fold of wild-type KBS and also enables a similar protein and DNA conformation and Y536 interactions as in Kaiso-KBS (Figure 5B). A similar steric clash in the Kaiso-MeCG2 complex is alleviated by a shift of the T7 base and DNA backbone away from Y536 and the α -helix of zinc finger 2, breaking the hydrogen bond between the Y536 hydroxyl and the T7 phosphate and leading to conformational disorder in the 5' region of the DNA. These findings strongly suggest that, in Kaiso complexes with KBS, the Y536 and Y562 interactions are critical for stabilizing the protein-DNA complex as well as promoting a specific DNA conformation, which could optimize the interactions at the binding interface.

By contrast, changes in DNA sequence and conformation that impair binding of Kaiso to KBS sites by almost 100-fold have much smaller effects (< 6 -fold) on binding to methylated DNA (Table 1). We observe only small differences in affinity with substitutions in the flanking sequence 5' to the mCpG site, which can be directly linked to the "in" conformation of E535. This is supported by data for the E535Q mutation, which mimics the E535 "in" conformation and leads to much smaller changes in affinity between KBS and its 5' variants compared to the wild-type or E535A protein. Selectivity for the 5' flanking base pair is further reduced in Kaiso-MeDNA complexes, where direct E535-mCpG contacts are strengthened by optimal CH---O hydrogen bonds and favorable electrostatics between the negatively charged glutamate and electropositive mC base. Kaiso is also able to accommodate different DNA conformations with methylated sites with only small changes in binding affinity. Thus, the DNA helix in the Kaiso-MeCG2 complex is deformed to relieve the Y536-T7 steric clash and displays two major conformations, one with reduced bending at the 5' flanking sequence, while the MeEcad complex also displays a different and less pronounced DNA helical bending.⁸⁷ As a result, interactions between Y536 and the phosphate backbone, which stabilize the KBS DNA conformation, can be disrupted without significant losses in binding affinity for methylated DNA. The Y536A mutation that substantially weakens Kaiso binding to KBS has little to no effect on binding to MeKBS or other methylated sites (Table 1). K539, on the other hand, adapts to the sequence changes and forms direct hydrogen bonds with specific bases (T7 in MeCG2) or the DNA backbone (A6 in MeEcad) as opposed to weaker, water-mediated ones in MeKBS (Figure 5). Diverse DNA helical conformations are observed in Kaiso-MeDNA complexes, in contrast to Kaiso-KBS complexes where nearly identical DNA conformations are observed in all of the crystal structures. Together, these data are consistent with a mechanism in which Kaiso binding to methylated DNA is dominated by direct readout of the 5' mCpG site, while direct (sequence) and indirect (shape) readout of the 5' flanking site play a secondary role.

We speculate that intrinsic variations in DNA conformation and flexibility at TpG (KBS) versus mCpG sites may contribute to differences in DNA readout by Kaiso. We observe that E535A, in contrast to wild-type Kaiso, has an order of magnitude higher affinity for KBS over MeKBS ($\Delta G \sim 1.3$ kcal/mol), showing a clear preference for binding to TpG over mCpG sites. Indeed, TpG/CpA steps are among the most flexible and conformationally heterogeneous dinucleotide steps and are commonly found at sites of helical deformation (kinking and bending) in protein-DNA complexes.^{60,62,65,109} NMR and molecular dynamics studies have shown that TpG/CpA steps display enhanced backbone mobility (sugar re-puckering and BI/BII transitions)^{57,60} and pronounced excursions to non-canonical Hoogsteen base pairs relative to other dinucleotide steps,¹¹⁰ which could be attributed to decreased stability of the Watson-Crick state.¹¹¹ Therefore, TpG sites may exhibit a higher propensity to deform into an optimal conformation for Kaiso binding that favors indirect readout. On the contrary, methylated CpG sites are known to rigidify the DNA helix and decrease its tendency to circularize and assemble into nucleosomes.⁷⁴ Cytosine methylation perturbs the local DNA shape,¹¹² hydration,¹¹³ and dynamics,⁵⁷ leading to changes in the amplitude and direction of motions. Since methylated CpG sites are less flexible and can reduce the flexibility of adjacent base pairs,¹¹³ methylated DNAs may be more resistant to deformation and incur a greater energetic cost to accommodate the Kaiso ZF domains. Thus,

it is likely that the higher inherent flexibility of TpG over mCpG sites underlies the preference of Kaiso E535A for KBS versus MeKBS motifs.

It has been previously hypothesized that CpG methylation can alter the indirect readout of DNA (i.e. conformation and dynamics) by transcription factors and nucleases. Methylation-induced changes in DNA backbone mobility have been implicated in the lower affinity of HhaI endonuclease for methylated versus unmethylated CpG sites.⁵⁷ Moreover, changes in DNA minor groove shape due to CpG methylation have been proposed to reduce cooperative binding by the Pbx-Hox transcription factor pair¹¹² or increase cleavage efficiency by the widely used footprinting enzyme DNaseI.¹¹⁴ Here, we provide evidence that, upon substitution of the critical E535 side chain, the lower intrinsic flexibility of mCpG sites has a large negative impact on the affinity of Kaiso for methylated versus non-methylated KBS DNA. In the case of Kaiso, strong interactions between E535 and the two methylated cytosines in the 5' mCpG site and additional contacts with the 3' mCpG site can compensate for lower DNA deformability and may also offset the negative impact of suboptimal DNA conformation and interactions at the 5' flanking site. It is conceivable that other MBPs use a similar strategy to maintain high affinity for methylated CpG sites and low sensitivity to their context in the DNA sequence. It may be especially true for zinc finger MBPs that commonly use acidic amino acids (glutamate or aspartate) to recognize 5-methylcytosine modifications,⁸ where favorable pairing between the negatively charged amino acid and the electropositive mC nucleobase, reinforced by hydrophobic and hydrogen bonding interactions (as in Kaiso), could select for 5-methylcytosine even when DNA conformation and dynamics are not optimal.

Conclusions

Various mechanisms have been proposed to explain the recognition of specific and methylated sequences by Kaiso⁸⁸ and other MBPs.⁴² Our studies provide the first direct insights into the molecular basis of differential readout of specific and methylated DNA motifs. Optimal binding of Kaiso to KBS sites requires a specific DNA sequence on the 5' flank of the TpG motif and depends on key interactions at the interface that stabilize an optimal DNA conformation. By contrast, Kaiso binding to methylated DNA is contingent on direct interactions between the carboxyl group of E535 and the 5' mCpG motif, while DNA conformation and sequence constitutes a less critical component of protein recognition. Our results suggest that the strong methyl-specific interactions between Kaiso and methylated DNA act to offset the requirement for specific flanking sequences, allowing for relatively indiscriminate binding to mCpG sites in regulatory regions and thus a more general ability to affect gene expression. Our findings also demonstrate that TFs can sense intrinsic variations in DNA shape and flexibility and adapt conformationally to achieve distinct DNA readout outcomes and target a greater variety of regulatory and epigenetic sites in the cell.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Gerard Kroon for help with NMR experiments, Maria Martinez-Yamout for valuable discussions, and Euvel Manlapaz for technical assistance. This work was supported by grant GM36643 from the National Institutes of Health (PEW) and by the Skaggs Institute for Chemical Biology. ENN was supported by a Robert Black fellowship from the Damon Runyon Cancer Research Foundation. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC0206CH11357. The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169.

References

- (1). Lee TI, and Young RA (2013) Transcriptional regulation and its misregulation in disease, *Cell* 152, 1237–1251. [PubMed: 23498934]
- (2). Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, and Weirauch MT (2018) The human transcription factors, *Cell* 175, 598–599. [PubMed: 30290144]
- (3). Latchman DS (1997) Transcription factors: an overview, *Int. J. Biochem. Cell Biol* 29, 1305–1312. [PubMed: 9570129]
- (4). Todeschini AL, Georges A, and Veitia RA (2014) Transcription factors: specific DNA binding and specific gene regulation, *Trends Genet.* 30, 211–219. [PubMed: 24774859]
- (5). Inukai S, Kock KH, and Bulyk ML (2017) Transcription factor-DNA binding: beyond binding site motifs, *Curr. Opin. Genet. Devel* 43, 110–119. [PubMed: 28359978]
- (6). Zhu H, Wang G, and Qian J (2016) Transcription factors as readers and effectors of DNA methylation, *Nat. Rev. Genet* 17, 551–565. [PubMed: 27479905]
- (7). Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schübeler D, Vinson C, and Taipale J (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors, *Science* 356, 502 (eaaj2239, pp1–15).
- (8). Hudson NO, and Buck-Koehntop BA (2018) Zinc finger readers of methylated DNA, *Molecules* 23, E2555 (pp1–22).
- (9). Breiling A, and Lyko F (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond, *Epigenet. Chromatin* 8, 24 (pp1–9).
- (10). Kim JK, Samaranayake M, and Pradhan S (2009) Epigenetic mechanisms in mammals, *Cell Mol. Life Sci* 66, 596–612. [PubMed: 18985277]
- (11). Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development, *Nat. Rev. Genet* 3, 662–673. [PubMed: 12209141]
- (12). Chen Z, Li S, Subramaniam S, Shyy JY, and Chien S (2017) Epigenetic regulation: a new frontier for biomedical engineers, *Annu. Rev. Biomed. Eng* 19, 195–219. [PubMed: 28301736]
- (13). Bird A (2002) DNA methylation patterns and epigenetic memory, *Genes Devel.* 16, 6–21. [PubMed: 11782440]
- (14). Li E, and Zhang Y (2014) DNA methylation in mammals, *Cold Spring Harb. Perspect. Biol* 6, a019133 (pp1–21).
- (15). Bird AP (1986) CpG-rich islands and the function of DNA methylation, *Nature* 321, 209–213. [PubMed: 2423876]
- (16). Klose RJ, and Bird AP (2006) Genomic DNA methylation: the mark and its mediators, *Trends Biochem. Sci* 31, 89–97. [PubMed: 16403636]

- (17). Lister R, and Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation, *Genome Res.* 19, 959–966. [PubMed: 19273618]
- (18). Keshet I, Lieman-Hurwitz J, and Cedar H (1986) DNA methylation affects the formation of active chromatin, *Cell* 44, 535–543. [PubMed: 3456276]
- (19). Kass SU, Pruss D, and Wolffe AP (1997) How does DNA methylation repress transcription?, *Trends Genet.* 13, 444–449. [PubMed: 9385841]
- (20). Fournier A, Sasai N, Nakao M, and Defossez PA (2012) The role of methyl-binding proteins in chromatin organization and epigenome maintenance, *Brief. Func. Genom* 11, 251–264.
- (21). Choy JS, Wei S, Lee JY, Tan S, Chu S, and Lee TH (2010) DNA methylation increases nucleosome compaction and rigidity, *J. Am. Chem. Soc* 132, 1782–1783. [PubMed: 20095602]
- (22). Bartolomei MS, and Ferguson-Smith AC (2011) Mammalian genomic imprinting, *Cold Spring Harb. Perspect. Biol* 3, a002592 (pp1–17).
- (23). Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, and Antonarakis SE (2011) DNA methylation profiles of human active and inactive X chromosomes, *Genome Res.* 21, 1592–1600. [PubMed: 21862626]
- (24). Robertson KD (2005) DNA methylation and human disease, *Nat. Rev. Genet* 6, 597–610. [PubMed: 16136652]
- (25). Klutstein M, Nejman D, Greenfield R, and Cedar H (2016) DNA methylation in cancer and aging, *Cancer Res.* 76, 3446–3450. [PubMed: 27256564]
- (26). Kulis M, and Esteller M (2010) DNA methylation and cancer, *Adv. Genet* 70, 27–56. [PubMed: 20920744]
- (27). Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, and Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*, *Cell* 126, 1189–1201. [PubMed: 16949657]
- (28). Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, Li Q, Luo R, Yu C, Yu J, Sun J, Zou X, Cao X, Xie X, Wang J, and Wang W (2012) Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression, *BMC Genom.* 13, 300 (pp1–15).
- (29). Laurent L, Wong E, Li G, Huynh T, Tsigiris A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, and Wei CL (2010) Dynamic changes in the human methylome during differentiation, *Genome Res.* 20, 320–331. [PubMed: 20133333]
- (30). Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, and Oliviero S (2017) Intragenic DNA methylation prevents spurious transcription initiation, *Nature* 543, 72–77. [PubMed: 28225755]
- (31). Bogdanovic O, and Veenstra G (2009) DNA methylation and methyl-CpG binding proteins: developmental requirements and function, *Chromosoma* 118, 549–565. [PubMed: 19506892]
- (32). Hendrich B, and Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins, *Mol. Cell. Biol* 18, 6538–6547. [PubMed: 9774669]
- (33). Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, and Defossez PA (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription, *Mol. Cell. Biol* 26, 169–181. [PubMed: 16354688]
- (34). Clouaire T, and Stancheva I (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin?, *Cell. Mol. Life Sci* 65, 1509–1522. [PubMed: 18322651]
- (35). Defossez PA, and Stancheva I (2011) Biological functions of methyl-CpG-binding proteins, *Prog. Mol. Biol. Transl. Sci* 101, 377–398. [PubMed: 21507359]
- (36). Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, and Rohs R (2015) Evolving insights on how cytosine methylation affects protein-DNA binding, *Brief. Func. Genom* 14, 61–73.
- (37). Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, and Prokhortchouk E (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor, *Genes Devel.* 15, 1613–1618. [PubMed: 11445535]

- (38). Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, Fitzgerald P, and Vinson C (2010) CpG methylation of half-CRE sequences creates C/EBP α binding sites that activate some tissue-specific genes, *Proc. Natl. Acad. Sci. U.S.A* 107, 20311–20316. [PubMed: 21059933]
- (39). Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone P, Grimaldi G, Riccio A, and Trono D (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions, *Mol. Cell* 44, 361–372. [PubMed: 22055183]
- (40). Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, and Cheng X (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence, *Genes Devel.* 28, 2304–2313. [PubMed: 25258363]
- (41). Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, and Cheng X (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA, *Mol. Cell* 66, 711–720. [PubMed: 28529057]
- (42). Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, Xia S, Liu S, Lyu H, Ming GL, Wade H, Song H, Qian J, and Zhu H (2013) DNA methylation presents distinct binding sites for human transcription factors, *Elife* 2, e00726 (pp1–16).
- (43). Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Muller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, and Vermeulen M (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell* 152, 1146–1159. [PubMed: 23434322]
- (44). Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR, and Vinson C (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo, *Genome Res.* 23, 988–997. [PubMed: 23590861]
- (45). Daniel JM, Spring CM, Crawford HC, Reynolds AB, and Baig A (2002) The p120^{cas}-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides, *Nucl. Acids Res* 30, 2911–2919. [PubMed: 12087177]
- (46). Rohs R, West SM, Sosinsky A, Liu P, Mann RS, and Honig B (2009) The role of DNA shape in protein-DNA recognition, *Nature* 461, 1248–1253. [PubMed: 19865164]
- (47). Rohs R, Jin X, West SM, Joshi R, Honig B, and Mann RS (2010) Origins of specificity in protein-DNA recognition, *Annu. Rev. Biochem* 79, 233–269. [PubMed: 20334529]
- (48). Kitayner M, Rozenberg H, Rohs R, Suad O, Rabinovich D, Honig B, and Shakked Z (2010) Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs, *Nat. Struct. Mol. Biol* 17, 423–429. [PubMed: 20364130]
- (49). Sasai N, and Defossez PA (2009) Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes, *Int. J. Dev. Biol* 53, 323–334. [PubMed: 19412889]
- (50). Liu Y, Zhang X, Blumenthal RM, and Cheng X (2013) A common mode of recognition for methylated CpG, *Trends Biochem. Sci* 38, 177–183. [PubMed: 23352388]
- (51). Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, and Cheng X (2014) Structural basis for Klf4 recognition of methylated DNA, *Nucl. Acids Res* 42, 4859–4867. [PubMed: 24520114]
- (52). Liu Y, Toh H, Sasaki H, Zhang X, and Cheng X (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence, *Genes Devel.* 26, 2374–2379. [PubMed: 23059534]
- (53). Nikolova EN, Stanfield RL, Dyson HJ, and Wright PE (2018) CH...O Hydrogen bonds mediate highly specific recognition of methylated CpG sites by the zinc finger protein Kaiso, *Biochemistry* 57, 2109–2120. [PubMed: 29546986]
- (54). Buck-Koehntop BA, Martinez-Yamout MA, Dyson HJ, and Wright PE (2012) Kaiso uses all three zinc fingers and adjacent sequence motifs for high affinity binding to sequence-specific and methyl-CpG DNA targets, *FEBS Lett.* 586, 734–739. [PubMed: 22300642]
- (55). Heinemann U, and Hahn M (1992) C-C-A-G-G-C-m5C-T-G-G. Helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G, *J. Biol. Chem* 267, 7332–7341. [PubMed: 1559976]

- (56). Schueler MG, and Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin, *Annu. Rev. Genomics Hum. Genet* 7, 301–313. [PubMed: 16756479]
- (57). Pederson K, Meints GA, Shajani Z, Miller PA, and Drobny GP (2008) Backbone dynamics in the DNA HhaI protein binding site, *J. Am. Chem. Soc* 130, 9072–9079. [PubMed: 18570423]
- (58). Echodu D, Goobes G, Shajani Z, Pederson K, Meints G, Varani G, and Drobny G (2008) Furanose dynamics in the HhaI methyltransferase target DNA studied by solution and solid-state NMR relaxation, *J. Phys. Chem. B* 112, 13934–13944. [PubMed: 18844399]
- (59). Tian Y, Kayatta M, Shultis K, Gonzalez A, Mueller LJ, and Hatcher ME (2009) ³¹P NMR investigation of backbone dynamics in DNA binding sites, *J. Phys. Chem. B* 113, 2596–2603. [PubMed: 18717548]
- (60). Nikolova EN, Bascom GD, Andricioaei I, and Al-Hashimi HM (2012) Probing sequence-specific DNA flexibility in A-tracts and pyrimidine-purine steps by nuclear magnetic resonance ¹³C relaxation and molecular dynamics simulations, *Biochemistry* 51, 8654–8664. [PubMed: 23035755]
- (61). Duchardt E, Nilsson L, and Schleucher J (2008) Cytosine ribose flexibility in DNA: a combined NMR ¹³C spin relaxation and molecular dynamics simulation study, *Nucl. Acids Res* 36, 4211–4219. [PubMed: 18579564]
- (62). Packer MJ, Dauncey MP, and Hunter CA (2000) Sequence-dependent DNA structure: dinucleotide conformational maps, *J. Mol. Biol* 295, 71–83. [PubMed: 10623509]
- (63). Packer MJ, Dauncey MP, and Hunter CA (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps, *J. Mol. Biol* 295, 85–103. [PubMed: 10623510]
- (64). Gardiner EJ, Hunter CA, Packer MJ, Palmer DS, and Willett P (2003) Sequence-dependent DNA structure: a database of octamer structural parameters, *J. Mol. Biol* 332, 1025–1035. [PubMed: 14499606]
- (65). Suzuki M, and Yagi N (1995) Stereochemical basis of DNA bending by transcription factors, *Nucl. Acids Res* 23, 2083–2091. [PubMed: 7610037]
- (66). Dickerson RE (1998) DNA bending: the prevalence of kinkiness and the virtues of normality, *Nucl. Acids Res* 26, 1906–1926. [PubMed: 9518483]
- (67). Dickerson RE, and Chiu TK (1997) Helix bending as a factor in protein/DNA recognition, *Biopolymers* 44, 361–403. [PubMed: 9782776]
- (68). Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, and Widom J (2006) A genomic code for nucleosome positioning, *Nature* 442, 772–778. [PubMed: 16862119]
- (69). Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, and Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast, *Nat. Genet* 39, 1235–1244. [PubMed: 17873876]
- (70). Meints GA, and Drobny GP (2001) Dynamic impact of methylation at the M. *HhaI* target site: a solid-state deuterium NMR study, *Biochemistry* 40, 12436–12443. [PubMed: 11591165]
- (71). Derreumaux S, Chaoui M, Tevanian G, and Fermandjian S (2001) Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element, *Nucl. Acids Res* 29, 2314–2326. [PubMed: 11376150]
- (72). Nathan D, and Crothers DM (2002) Bending and flexibility of methylated and unmethylated EcoRI DNA, *J. Mol. Biol* 316, 7–17. [PubMed: 11829499]
- (73). Mirsaidov U, Timp W, Zou X, Dimitrov V, Schulten K, Feinberg AP, and Timp G (2009) Nanoelectromechanics of methylated DNA in a synthetic nanopore, *Biophys. J* 96, L32–34. [PubMed: 19217843]
- (74). Perez A, Castellazzi CL, Battistini F, Collinet K, Flores O, Deniz O, Ruiz ML, Torrents D, Eritja R, Soler-Lopez M, and Orozco M (2012) Impact of methylation on the physical properties of DNA, *Biophys. J* 102, 2140–2148. [PubMed: 22824278]
- (75). Horberg J, and Reymer A (2018) A sequence environment modulates the impact of methylation on the torsional rigidity of DNA, *Chem. Comm* 54, 11885–11888. [PubMed: 30288518]
- (76). Ngo TT, Yoo J, Dai Q, Zhang Q, He C, Aksimentiev A, and Ha T (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability, *Nat. Commun* 7, 10813 (pp1–9).

- (77). Felle M, Hoffmeister H, Rothhammer J, Fuchs A, Exler JH, and Langst G (2011) Nucleosomes protect DNA from DNA methylation *in vivo* and *in vitro*, Nucl. Acids Res 39, 6956–6969. [PubMed: 21622955]
- (78). Daniel JM, and Reynolds AB (1999) The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor, Mol. Cell. Biol 19, 3614–3623. [PubMed: 10207085]
- (79). Park J-I, Kim SW, Lyons JP, Ji H, Nguyen TT, Cho K, Barton MC, Deroo T, Vleminckx K, and McCrea PD (2005) Kaiso/p120-catenin and TCF/ β -catenin complexes coordinately regulate canonical Wnt gene targets, Devel. Cell 8, 843–854. [PubMed: 15935774]
- (80). Spring CM, Kelly KF, O’Kelly I, Graham M, Crawford HC, and Daniel JM (2005) The catenin p120ctn inhibits Kaiso-mediated transcriptional repression of the β -catenin/TCF target gene matrilysin, Exp. Cell Res 305, 253–265. [PubMed: 15817151]
- (81). Ogden SR, Wroblewski LE, Weydig C, Romero-Gallo J, O’Brien DP, Israel DA, Krishna US, Fingleton B, Reynolds AB, Wessler S, and Peek RM Jr. (2008) p120 and Kaiso regulate *Helicobacter pylori*-induced expression of matrix metalloproteinase-7, Mol. Biol. Cell 19, 4110–4121. [PubMed: 18653469]
- (82). Donaldson NS, Pierre CC, Anstey MI, Robinson SC, Weerawardane SM, and Daniel JM (2012) Kaiso represses the cell cycle gene cyclin D1 via sequence-specific and methyl-CpG-dependent mechanisms, PLoS ONE 7, e50398 (pp1–12).
- (83). Pierre CC, Longo J, Bassey-Archibong BI, Hallett RM, Milosavljevic S, Beatty L, Hassell JA, and Daniel JM (2015) Methylation-dependent regulation of hypoxia inducible factor-1 alpha gene expression by the transcription factor Kaiso, Biochim. Biophys. Acta. Gene Regul. Mech 1849, 1432–1441.
- (84). Wang H, Liu W, Black S, Turner O, Daniel JM, Dean-Colomb W, He QP, Davis M, and Yates C (2016) Kaiso, a transcriptional repressor, promotes cell migration and invasion of prostate cancer cells through regulation of miR-31 expression, Oncotarget 7, 5677–5689. [PubMed: 26734997]
- (85). Yoon HG, Chan DW, Reynolds AB, Qin J, and Wong J (2003) N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso, Mol. Cell 12, 723–734. [PubMed: 14527417]
- (86). Raghav SK, Waszak SM, Krier I, Gubelmann C, Isakova A, Mikkelsen TS, and Deplancke B (2012) Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBP β and KAISO, Mol. Cell 46, 335–350. [PubMed: 22521691]
- (87). Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, and Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso, Proc. Natl. Acad. Sci. U.S.A 109, 15229–15234. [PubMed: 22949637]
- (88). Sasai N, Nakao M, and Defossez PA (2010) Sequence-specific recognition of methylated DNA by human zinc-finger proteins, Nucl. Acids Res 38, 5015–5022. [PubMed: 20403812]
- (89). Otwinowski Z, and Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode, In Methods Enzymol Macromol. Crystall. A (Carter CW Jr., and Sweet RM, Eds.), pp 307–326, Academic Press, New York.
- (90). McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007) Phaser crystallographic software, J. Appl. Crystallogr 40, 658–674. [PubMed: 19461840]
- (91). Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010) Features and development of Coot, Acta Cryst. D 66, 486–501. [PubMed: 20383002]
- (92). Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Cryst. D 66, 213–221. [PubMed: 20124702]
- (93). Schrödinger L (2013) The PyMOL Molecular Graphics System.
- (94). Li S, Olson WK, and Lu XJ (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures, Nucl. Acids Res 47, W26–W34. [PubMed: 31114927]

- (95). Delaglio F, Grzesiek S, Vuister GW, Guang Z, Pfeifer J, and Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6, 277–293. [PubMed: 8520220]
- (96). Goddard TD, and Kneller DG (2006) SPARKY 3, University of California, San Francisco.
- (97). Korzhnev DM, Orekhov VY, and Kay LE (2005) Off-Resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a Fyn SH3 domain, *J. Am. Chem. Soc* 127, 713–721. [PubMed: 15643897]
- (98). Hansen AL, Nikolova EN, Casiano-Negroni A, and Al-Hashimi HM (2009) Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy, *J. Am. Chem. Soc* 131, 3818–3819. [PubMed: 19243182]
- (99). Miloushev VZ, and Palmer III AG (2005) R1ρ relaxation for two-site chemical exchange: General approximations and some exact solutions, *J. Magn. Reson* 177, 221–227. [PubMed: 16143548]
- (100). Platzer G, Okon M, and McIntosh L (2014) pH-dependent random coil ¹H, ¹³C, and ¹⁵N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements, *J. Biomol. NMR* 60, 109–129. [PubMed: 25239571]
- (101). Jaremko Ł, Jaremko M, Ejchart A, and Nowakowski M (2018) Fast evaluation of protein dynamics from deficient 15N relaxation data, *J. Biomol. NMR* 70, 219–228. [PubMed: 29594733]
- (102). Mulder FA, De Graaf RA, Kaptein R, and Boelens R (1998) An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations, *J. Magn. Reson* 131, 351–357. [PubMed: 9571112]
- (103). Massi F, Johnson E, Wang C, Rance M, and Palmer AG III. (2004) NMR R_{1ρ} rotating-frame relaxation with weak radio frequency fields, *J. Am. Chem. Soc* 126, 2247–2256. [PubMed: 14971961]
- (104). Wallerstein J, Weininger U, Khan MAI, Linse S, and Akke M (2015) Site-specific protonation kinetics of acidic side chains in proteins determined by pH-dependent carboxyl ¹³C NMR relaxation, *J. Am. Chem. Soc* 137, 3093–3101. [PubMed: 25665463]
- (105). Deng H, Dyer RB, and Callender R (2019) Active-site Glu165 activation in triosephosphate isomerase and its deprotonation kinetics, *J. Phys. Chem. B* 123, 4230–4241. [PubMed: 31013084]
- (106). Dixit SB, Arora N, and Jayaram B (2000) How do hydrogen bonds contribute to protein-DNA recognition, *J. Biomol. Struct. Dynam* 17, 109–112.
- (107). Harris RC, Mackoy T, Dantas Machado AC, Xu D, Rohs R, and Fenley MO (2012) Chapter 3 Opposites Attract: Shape and Electrostatic Complementarity in Protein-DNA Complexes, In *Innovations in Biomolecular Modeling and Simulations: Volume 2*, pp 53–80, The Royal Society of Chemistry.
- (108). Liebl K, and Zacharias M (2019) How methyl-sugar interactions determine DNA structure and flexibility, *Nucl. Acids Res* 47, 1132–1140. [PubMed: 30541032]
- (109). Okonogi TM, Alley SC, Reese AW, Hopkins PB, and Robinson BH (2002) Sequence-dependent dynamics of duplex DNA: the applicability of a dinucleotide model, *Biophys. J* 83, 3446–3459. [PubMed: 12496111]
- (110). Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, and Al-Hashimi HM (2011) Transient Hoogsteen base pairs in canonical duplex DNA, *Nature* 470, 498–502. [PubMed: 21270796]
- (111). Alvey HS, Gottardo FL, Nikolova EN, and Al-Hashimi HM (2014) Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics, *Nat. Commun* 5, 4786 (pp1–8).
- (112). Rao S, Chiu TP, Kribelbauer JF, Mann RS, Bussemaker HJ, and Rohs R (2018) Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding, *Epigenetics Chromatin* 11, 6 (pp1–11). [PubMed: 29310712]
- (113). Teng X, and Hwang W (2018) Effect of Methylation on Local Mechanics and Hydration Structure of DNA, *Biophys. J* 114, 1791–1803. [PubMed: 29694859]

- (114). Lazarovici A, Zhou T, Shafer A, Dantas Machado AC, Riley TR, Sandstrom R, Sabo PJ, Lu Y, Rohs R, Stamatoyannopoulos JA, and Bussemaker HJ (2013) Probing DNA shape and methylation state on a genomic scale with DNase I, Proc. Natl. Acad. Sci. U.S.A 110, 6376–6381. [PubMed: 23576721]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Solution conformation and dynamics of E535 in Kaiso bound to KBS. (A) Crystal structures of Kaiso bound to KBS (PDB 4F6M⁸⁷) (green) and to a methylated DNA with a sequence similar to KBS (PDB 5VMW⁵³) (MeKBS, salmon). Canonical (colored) and CH---O (black) hydrogen bonds between E535 and the TpG (KBS) or mCpG (MeKBS) site are displayed with dashed lines and distances are labeled in Å. (B) pH titration curves for E535 and Y536 in Kaiso-KBS and best fits (blue line) to a two-state equilibrium model. (C) Overlay of portions of the backbone amide ^1H , ^{15}N -HSQC spectra for Kaiso-KBS at pH 7.0 (grey) or pH 4.8 (purple), Kaiso(E535Q)-KBS (green), Kaiso-MeKBS (salmon), Kaiso-MeEcad (blue), and Kaiso-MeCG2 (yellow) (all at pH 7.0) showing the chemical shifts of E535 and Y536 (side chain amides labeled with asterisks). Dashed ovals show the position of E535 and Y536 cross peaks in Kaiso-MeDNA complexes. Arrows indicate the ^{15}N chemical shifts of E535 and Y536 in the minor state of Kaiso-KBS at pH 7.0 obtained from $R_{1\rho}$ relaxation dispersion. (D) ^{15}N $R_{1\rho}$ relaxation dispersion profiles for E535 and Y536 showing $R_{1\rho}$ ($= R_2 + R_{\text{ex}}$) as a function of spin-lock power (ω_{SL} , on-resonance) and resonance offset from the spin lock (Ω). Solid lines indicate best fits to a two-state chemical exchange model and errors represent one s. d.

Figure 2.

Conformational switch at E535 in wild-type (WT) and mutant Kaiso complexes with KBS. Crystal structures of (A) Kaiso-KBS (pH 7.2), (B) Kaiso-KBS (pH 6.5), (C) Kaiso(K539A)-KBS (pH 7.2), (E) Kaiso(E535Q)-KBS (pH 7.2), and (F) Kaiso(E535A)-KBS (pH 7.2) showing the local conformation and interactions of E535 and K539. Direct or water-mediated hydrogen bonds are displayed with colored dashed lines, and potential CH---O hydrogen bonds with the T8 methyl group (magenta sphere) are indicated in black. (D) Overlay of ^1H , ^{15}N -HSQC spectra for WT Kaiso and K539A mutant bound to KBS or methylated DNA (MeCG2) showing cross peaks of E535 in the "out" or "in" (dotted circles) conformation. The E535 cross peak for Kaiso(K539A)-KBS is broadened beyond detection at pH 7.0, but the protonated E535 state in Kaiso(K539A)-KBS at pH 5.4 (red) is observed in the vicinity of E535 in the WT-KBS complex at pH 4.8 (purple). The E535 cross peaks in spectra of the WT-MeCG2 (yellow) and Kaiso(K539A)-MeCG2 (blue) complexes are observable at pH 7.0 at similar ^1H and ^{15}N chemical shifts.

Figure 4.

Conformation of Kaiso K539 in complex with different DNA sequences. Previously published crystal structures of (A) Kaiso-MeKBS at pH 7.2 (PDB ID: 5VMW), (B) Kaiso-MeCG2 at pH 8.5 (PDB ID: 5VMV), and (C) Kaiso-MeEcad at pH 6.5 (PDB ID: 4F6N) showing distinct conformations and DNA interactions at K539 depending on the sequence 5' to the mCpG site. (D) Current crystal structure of Kaiso-KBS(5TCG) at pH 5.0, showing a direct hydrogen bond between K539 and the 5' flanking G-C base pair.

Figure 5.

Role of DNA conformation in binding of WT Kaiso to specific and methylated DNA. (A) Interactions of Y536 and Y562 with the 5' flanking sequence in MeCG2 (gold) and KBS (grey) complexes, showing that a 5' T-A base pair (T7-A30) in MeCG2 displaces the DNA helix relative to KBS (5' C-G) and disrupts direct hydrogen bond and hydrophobic contacts of Y536 with the 5' T7 phosphate and C6 sugar moiety. (B) Molecular model of a 5' T-A (green, KBS(5TCT)) and A-T (blue, KBS(5TCA)) base pair in a Kaiso complex overlaid with the 5' C-G base pair in the Kaiso-KBS crystal structure, showing steric clash between the T methyl group and Y536 and Y562, respectively. An overlay of the crystal structures of the Kaiso-KBS and Kaiso-KBS(5TCU/A) complexes in the vicinity of the 5' C-G and 5' U-A base pair shows that the conformations are nearly identical after removal of the T methyl group. (C) Overlay of Kaiso-KBS (pH 7.2) and Kaiso-MeCG2 structures showing a large DNA deformation and differences in DNA conformation at the 5' flanking site between the two complexes. Key residues that directly interact with the 5' C-G (KBS) and 5' T-A (MeCG2) base pair (Y536 and Y562) and the 5' neighboring sequence (H543 and Q563) are highlighted in blue for the Kaiso-KBS and magenta for the Kaiso-MeCG2 structures. (D)

Affinity of WT Kaiso and E535Q and E535A mutants for KBS(5TCU/A) and KBS(5TCA/U) relative to KBS (expressed as a K_d ratio).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Kaiso-DNA binding kinetics and affinities.

DNA Construct	$k_{off} (s^{-1}) \times (10^{-4})$	$k_{on} (Ms^{-1})^b \times (10^4)$	$K_d (nM)$	$k_{off} (s^{-1})^a \times (10^{-4})$	$k_{on} (Ms^{-1})^b \times (10^4)$	$K_d (nM)$	$k_{off} (s^{-1})^a \times (10^{-4})$	$k_{on} (Ms^{-1})^b \times (10^4)$	$K_d (nM)$
Kaiso WT									
KBS	1.3 ± 0.1	4.6 ± 0.5	2.8 ± 0.4	0.23 ± 0.02	4.3 ± 0.5	0.53 ± 0.07	2.0 ± 0.2	4.9 ± 0.7	4.0 ± 0.7
KBS(5TCG)	26 ± 1	4.6 ± 0.3	58 ± 5	0.18 ± 0.01	5.4 ± 0.4	0.33 ± 0.03	9.0 ± 0.2	4.1 ± 0.2	22 ± 1
KBS(5TCT)	95 ± 4	d	(210 ± 20)	2.6 ± 0.1	4.5 ± 0.1	5.8 ± 0.3	49 ± 2	d	(110 ± 10)
KBS(5TCA)	104 ± 11	d	(230 ± 30)	8.5 ± 0.3	4.5 ± 0.1	19 ± 1	53 ± 6	d	(120 ± 20)
KBS(5TCU/A)	5.0 ± 0.2	4.0 ± 0.7	12 ± 2	0.42 ± 0.04	4.3 ± 0.4	1.0 ± 0.1	2.9 ± 0.5	4.1 ± 0.2	7.1 ± 1.2
KBS(5TCA/U)	9.5 ± 0.9	4.1 ± 0.6	23 ± 4	0.63 ± 0.03	5.4 ± 0.2	1.2 ± 0.1	14 ± 1	4.3 ± 0.6	31 ± 5
MeKBS ^c	0.13 ± 0.02	6.4 ± 0.5	0.20 ± 0.04	3.9 ± 0.4	5.4 ± 0.4	7.2 ± 0.9	16 ± 2	d	(35 ± 5)
MeCG2 ^c	0.23 ± 0.02	5.3 ± 0.6	0.43 ± 0.06	8.4 ± 0.4	4.2 ± 0.2	20 ± 1	34 ± 2	d	(76 ± 9)
MeCG2B	0.53 ± 0.02	4.5 ± 0.4	1.2 ± 0.1	19 ± 1	5.1 ± 0.2	38 ± 1	47 ± 1	d	(100 ± 10)
MeEcad	0.34 ± 0.04	3.9 ± 0.4	0.87 ± 0.13	21 ± 1	5.3 ± 0.6	40 ± 1	48 ± 1	d	(110 ± 10)
Kaiso Y536A									
KBS	150 ± 10	d	(340 ± 40)	300 ± 30	d	(670 ± 100)	4.1 ± 0.4	4.4 ± 1.0	9.3 ± 2.2
KBS(5TCG)	380 ± 20	d	(840 ± 90)	340 ± 40	d	(760 ± 120)	190 ± 30	5.2 ± 1.0	370 ± 90
KBS(5TCT)	120 ± 10	d	(270 ± 30)	320 ± 50	d	(710 ± 130)	190 ± 30	d	(410 ± 80)
KBS(5TCA)	280 ± 20	d	(630 ± 80)	360 ± 10	d	(790 ± 80)	35 ± 4	4.2 ± 1.0	83 ± 9
KBS(5TCU/A)	150 ± 10	d	(340 ± 40)	400 ± 20	d	(880 ± 10)	250 ± 40	d	(560 ± 60)
KBS(5TCA/U)	240 ± 20	d	(540 ± 70)	390 ± 70	d	(860 ± 90)	60 ± 5	4.5 ± 0.7	130 ± 20
MeKBS	0.26 ± 0.01	9.0 ± 0.5	0.29 ± 0.02	n/d^e	n/d^e	n/d^e	n/d^e	n/d^e	n/d^e
MeCG2	0.28 ± 0.02	10.0 ± 0.7	0.28 ± 0.03	5.9 ± 0.9	5.5 ± 0.7	11 ± 2	1.2 ± 0.1	8.8 ± 0.3	1.4 ± 0.1
MeCG2B	3.1 ± 0.2	7.5 ± 0.8	4.1 ± 0.5	8.3 ± 0.5	6.5 ± 1.3	22 ± 4	14 ± 1	7.5 ± 1.1	19 ± 3
MeEcad	1.5 ± 0.2	7.7 ± 0.8	2.0 ± 0.2	16 ± 1	7.2 ± 1.4	13 ± 3	1.5 ± 0.1	8.3 ± 1.3	1.8 ± 0.3

^a k_{off} values represent the mean and standard deviation of four or more individual kinetics measurements.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

k_{on} values represent the mean and standard deviation of two or more linear fits of the observed association rate constant as a function of protein concentration (also see Experimental section).

c values previously reported using the same assay and conditions.

d k_{on} could not be reliably determined for values in parentheses (see Materials and Methods). Approximate K_d values (shown in parentheses) were estimated using the measured k_{off} and the average value of k_{on} ($4.5 \pm 0.5 \times 10^4 \text{ Ms}^{-1}$) determined for binding of Kaiso to other DNA sequences.

e not determined