

# Structural basis of water-mediated *cis* Watson–Crick/Hoogsteen base-pair formation in non-CpG methylation

Shan-Meng Lin<sup>1</sup>, Hsiang-Ti Huang<sup>1</sup>, Pei-Ju Fang<sup>2</sup>, Chi-Fon Chang<sup>2</sup>, Roshan Satange<sup>1b</sup>, Chung-ke Chang<sup>3</sup>, Shan-Ho Chou<sup>4</sup>, Stephen Neidle<sup>5,\*</sup> and Ming-Hon Hou<sup>1,6,7,8,9,\*</sup>

<sup>1</sup>Graduate Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 402, Taiwan

<sup>2</sup>Genomics Research Center, Academia Sinica, Taipei 115, Taiwan

<sup>3</sup>Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan

<sup>4</sup>Institute of Biochemistry, National Chung Hsing University, Taichung 402, Taiwan

<sup>5</sup>School of Pharmacy, University College London, London WC1N 1AX, UK

<sup>6</sup>Doctoral Program in Medical Biotechnology, National Chung Hsing University, Taichung 402, Taiwan

<sup>7</sup>Graduate Institute of Biotechnology, National Chung Hsing University, Taichung 402, Taiwan

<sup>8</sup>Department of Life Sciences, National Chung Hsing University, Taichung 402, Taiwan

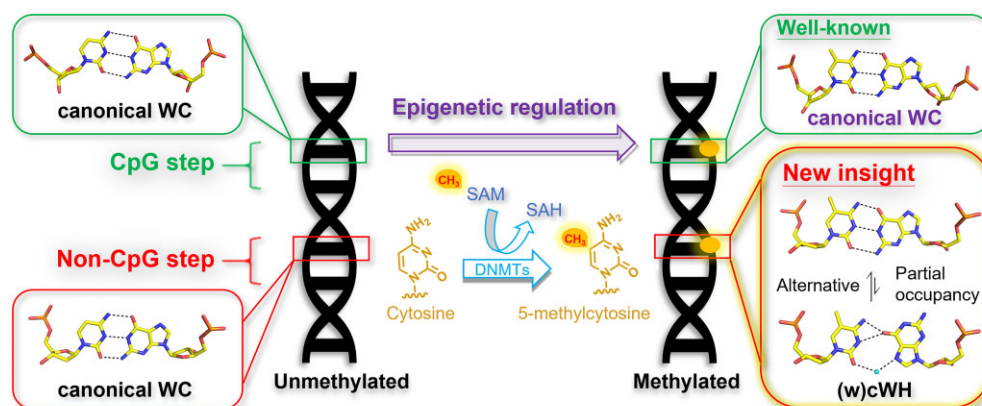
<sup>9</sup>Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

\*To whom correspondence should be addressed. Tel: +886 4 2284 0338 (Ext. 7011); Fax: +886 4 2285 9329; Email: mhho@nchu.edu.tw  
Correspondence may also be addressed to Stephen Neidle. Tel: +886 4 2284 0338 (Ext. 7011); Email: s.neidle@ucl.ac.uk

## Abstract

Non-CpG methylation is associated with several cellular processes, especially neuronal development and cancer, while its effect on DNA structure remains unclear. We have determined the crystal structures of DNA duplexes containing -CGCCG- regions as CCG repeat motifs that comprise a non-CpG site with or without cytosine methylation. Crystal structure analyses have revealed that the <sub>m</sub>C:G base-pair can simultaneously form two alternative conformations arising from non-CpG methylation, including a unique water-mediated *cis* Watson–Crick/Hoogsteen, (w)cWH, and Watson–Crick (WC) geometries, with partial occupancies of 0.1 and 0.9, respectively. NMR studies showed that an alternative conformation of methylated <sub>m</sub>C:G base-pair at non-CpG step exhibits characteristics of cWH with a *syn*-guanosine conformation in solution. DNA duplexes complexed with the DNA binding drug echinomycin result in increased occupancy of the (w)cWH geometry in the methylated base-pair (from 0.1 to 0.3). Our structural results demonstrated that cytosine methylation at a non-CpG step leads to an *anti*→*syn* transition of its complementary guanosine residue toward the (w)cWH geometry as a partial population of WC, in both drug-bound and naked <sub>m</sub>C:G base pairs. This particular geometry is specific to non-CpG methylated dinucleotide sites in B-form DNA. Overall, the current study provides new insights into DNA conformation during epigenetic regulation.

## Graphical abstract



## Introduction

DNA methylation is an important epigenetic modification mechanism that enables the regulation of gene expression

without altering its DNA sequence (1–5). In this process a methyl group is transferred to the pyrimidine ring of a cytosine at the fifth carbon position, forming a stable 5-methyl

Received: January 25, 2024. Revised: May 30, 2024. Editorial Decision: June 21, 2024. Accepted: June 28, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

cytosine ( $5_mC$ ) residue. The reaction is catalyzed by specific enzymes termed DNA methyltransferases and is tightly regulated. Aberrant methylation is a well-established molecular lesion in cancer (6–15) and has also been associated with other diseases, including neurodegeneration (2,8,16,17). There are two types of DNA methylation: CpG methylation occurs at cytosine residues immediately preceding a guanine residue and is often found in promoter regions (18,19). The process is catalyzed by the DNMT1 methyltransferase. On the other hand, methylation of cytosines immediately preceding a non-guanine residue is known as non-CpG methylation (15,20–27). Non-CpG methylation is often observed within genes and is catalyzed by another set of DNA methyltransferases, DNMT3a and DNMT3b (28–30). Other proteins, including the Enhancer of Zeste Homolog 2 (EZH2) and the methyl-CpG binding protein 2 (MeCP2), are also involved in the regulation of non-CpG methylation (31–35). Interestingly, non-CpG methylation in the human genome has been shown to correlate spatially with CpG methylation, suggesting a functional connection between the two types of DNA methylation (21).

Methylation of DNA regulates gene expression via two possible mechanisms: one is through the direct effect of the methyl moiety, which may physically interact with DNA-binding proteins and either enhance or inhibit their recruitment. The other is by exerting an indirect effect on the biophysical properties of DNA, such as changing its conformation or rigidity (36–43). Studying the effects of DNA methylation on DNA structure is thus crucial for understanding gene regulation. However, most structural studies have focused on the various aspects of CpG methylation. For non-CpG methylation, studies have focused on the complexes formed between the methylated DNA and its binding protein (44–50), but relatively little is known about the structural effects of non-CpG methylation on DNA itself.

In DNA structure, Watson–Crick (WC) base pairs, characterized by their specific hydrogen bonding patterns between complementary nucleobases (A–T, G–C), play a major role in stabilizing the DNA double helix, which is essential for the storage and replication of genetic information (51). Additionally, nucleotide base pairs can also adopt a variety of geometries under specific conditions. These base pairs, commonly referred to as non-canonical base pairs (52), have their distinctive hydrogen bonding patterns differing from those observed in canonical WC and have been well-defined (53). Non-canonical base pairs can participate in those biological processes that cannot be achieved through WC pairing alone, thereby broadening the structural and functional diversity of the DNA duplex (54–56). For example, the Hoogsteen (HG) base pair (57), containing a purine nucleotide undergoes a 180° rotation around the glycosidic bond (58), and can play a pivotal role in the specific binding of some proteins to DNA (59). Accordingly, it has been reported to be involved in a range of biological processes (60,61), such as DNA replication (61,62), DNA repair (60), as well as DNA recognition by the tumor suppressor p53 (63–65). On the other hand, recent NMR relaxation experiments on native DNA have revealed that the predominant WC base pairing is in dynamic equilibrium with a transient HG (66–70), and this WC-to-HG transition may be crucial for protein recognition (61,67,71). Understanding the geometric differences of DNA base pairs under specific biological conditions is thus of considerable potential

significance, as it may provide insights into the mechanisms governing genetic processes beyond the primary information encoded within the nucleotide sequence.

To investigate the structural effects of non-CpG methylation on the DNA structure, we have determined the crystal structures of the DNA duplexes d(ACG<sub>m</sub>CCGT/ACGGCGT) (named as <sub>m</sub>C:G pairing structure) and d(ACGCCGT/ACGGCGT) (named as C:G pairing structure), which represent methylated and unmethylated states, respectively. These crystal structures have revealed that cytosine methylation at a non-CpG site flanking two CpGs resulted in polymorphic conformations of the central <sub>m</sub>C:G base pair, including the unexpected formation of a novel water-mediated *cis* Watson–Crick/Hoogsteen, (w)cWH, base pair arrangement at partial occupancy, which is specific to methylation at a non-CpG site in B-form DNA. Consistent with these structural findings, parallel NMR studies have indicated that non-CpG methylated <sub>m</sub>C:G base-pair can adopt two alternative conformations in solution, one of which exhibited characteristics of Hoogsteen geometry. Additionally, we applied a DNA binding drug, echinomycin (ECHI), which is expected to stabilize (w)cWH near the drug intercalating site, to manipulate the occurrence of (w)cWH. An increased occupancy of (w)cWH geometry (from 0.1 to 0.3) was observed in the crystal structure of the <sub>m</sub>C:G-ECHI complex. This suggests that a conserved base pair geometry could also form with a non-CpG methylated <sub>m</sub>C:G base pair upon drug binding. These results highlight the significance of the unique (w)cWH geometry and provide the first insight into the effects of non-CpG methylation on the structure of a DNA duplex.

## Materials and methods

### Oligonucleotide sample preparation

All chemical reagents were purchased from Sigma Chemical Co. (St. Louis, MO, USA). DNA oligonucleotides were purchased from MD Bio (Taipei, Taiwan). DNA and echinomycin stock solutions were prepared in water and DMSO, respectively. The DNA oligonucleotide and echinomycin concentrations were determined by measuring the optical density at 260 and 440 nm, respectively, using a JASCO V-630 spectrophotometer (JASCO International Co. Ltd., Tokyo, Japan). The C:G [d(ACGCCGT/ACGGCGT)], <sub>m</sub>C:G [d(ACG<sub>m</sub>CCGT/ACGGCGT)] and C:I [d(ACGCCGT/ACGICGT)] pairing duplexes were prepared in water via heating at 95°C for 5 min followed by cooling on ice for 30 min to allow duplex formation.

### Crystallization

To obtain the unliganded C:G pairing crystals, 0.5 mM of the C:G pairing duplex was mixed in a solution containing 50 mM MES (pH = 6.5), 1 mM spermine HCl, 20% PEG400, and equilibrated against 500 μl of 30% PEG 400 at 20°C for 1 week. To obtain the unliganded <sub>m</sub>C:G pairing crystals, 0.45 mM of <sub>m</sub>CG duplex was mixed in a solution containing 50 mM MES (pH = 6.5), 1 mM spermine HCl, 20% PEG400, and equilibrated against 500 μl of 30% PEG 400 at 20°C for 1 week. To obtain the unliganded C:I pairing crystals, 0.125 mM of the C:I pairing duplex was mixed in a solution containing 30 mM MES (pH = 6.5), 1 mM spermine HCl, 8%

PEG400, 1.5 mM MgCl<sub>2</sub>, and equilibrated against 500 µl of 50% PEG 400 at 20°C for 1 week. To prepare the DNA:ECHI complexes, 0.125 mM DNA duplex was pre-incubated with 0.275 mM ECHI at 4°C for 72 h. For the C:G-ECHI complex, crystals were grown for three weeks in a 5 µl drop containing 20 mM MES (pH = 6), 10 mM spermine HCl, 1% PEG 200, 10 mM MnCl<sub>2</sub>, equilibrated against 500 µl of 30% PEG 200 at 20°C. For the <sub>m</sub>C:G-ECHI complex, crystals were grown for three weeks in a 5 µl drop containing 20 mM MES (pH = 6), 5 mM spermine HCl, 1.2% PEG 200, 10 mM MnCl<sub>2</sub>, equilibrated against 500 µl of 30% PEG 200 at 20°C.

### X-ray data collection, phasing, and structure refinement

X-ray diffraction data of unliganded C:G, <sub>m</sub>C:G, C:I pairing crystals and C:G-ECHI, <sub>m</sub>C:G-ECHI complex crystals were collected at the National Synchrotron Radiation Research Center, Hsinchu, Taiwan. The diffraction data for the unliganded C:G, C:I pairing crystals and the C:G-ECHI complex crystals were collected at beamline BL15A1 and recorded using a Rayonix MX300HE CCD area detector. Data for unliganded <sub>m</sub>C:G pairing crystals and <sub>m</sub>C:G-ECHI complex crystals were collected on beamline TPS 05A with a Rayonix MX300HS CCD detector. Diffraction data integration and reduction were conducted using the HKL-2000 package (72). The phases of these structures were solved using phaser-MR in PHENIX v1.17.1. The phases of the unliganded duplexes and DNA:ECHI complexes were solved using the partial structure of the DNA duplex (PDB ID:6JV5) (73) and ECHI-d(ACGTCGT)<sub>2</sub> (PDB ID:5YTZ) (74), respectively. Structural refinements were performed using the PHENIX package (v1.17.1.). In brief, all crystal structures were refined using the PHENIX package through three cycles of refinement in phenix.refine with default settings, along with the inclusion of hydrogen bond restraints. The refinement programs generated 2mF<sub>o</sub>-DF<sub>c</sub> and mF<sub>o</sub>-DF<sub>c</sub> difference maps, where 'F<sub>o</sub> and F<sub>c</sub>' are the experimentally observed and calculated amplitudes, respectively. 'm' denotes the figure of merit, and 'D' represents the Sigma-A weighting factor (75). The 2mF<sub>o</sub>-DF<sub>c</sub> map indicates the area where the model is most expected to be located. The mF<sub>o</sub>-DF<sub>c</sub> map, depicted with positive and negative contours, reveals the areas where atoms are lacking in the current model or where atoms exist in the model but are not present in the crystal, respectively (76). The refinement of (w)cWH base pairs were completed followed the subsequent procedure: (i) <sub>m</sub>C:G base pairs in the <sub>m</sub>C:G pairing duplex and <sub>m</sub>C:G-ECHI complex structures were initially modeled with a single guanosine in either the *anti* or *syn* conformation, and refined using the aforementioned phenix.refine routine. Following refinement, ambiguous maps were observed at the methylated <sub>m</sub>C:G base pairs in one of the four asymmetric units of the unliganded <sub>m</sub>C:G pairing structure and in the <sub>m</sub>C:G-ECHI structures (see [Supplementary Figure S13A, B and E, F](#)). (ii) An alternative guanosine residue was modeled into <sub>m</sub>C:G pairs, and a second round of refinement was carried out with the same phenix.refine routine. (iii) Based on mF<sub>o</sub>-DF<sub>c</sub> difference maps depicted with positive contours, a water molecule was modelled into <sub>m</sub>C:G pair and further refined using the same phenix.refine routine (see [Supplementary Figure S13C and G](#)). The refinement was ultimately optimized through B-factor refinement and group occupancy refinement (see [Supplementary Figure S13D and H](#))

### MD simulation protocol

The MD simulation was conducted based on the protocol modified from previous studies (77). Briefly, crystal structures of the representative duplexes for unliganded C:G, <sub>m</sub>C:G, and C:I pairing structures were used for MD simulations, and Discovery Studio software v19.1.0.18287 with the CHARMM force field was applied for the simulations (78). After a 1000 step smart-minimizer, the models were immersed in a water box built with a modified TIP3P system (79). The total charge on each system was neutralized by the random addition of sodium ions. Then a 5000-step Steepest Descent (SD) and 10 000-step conjugate gradient minimization was run, followed by heating from 50 to 300 K for 20 ps with position constraints on the DNA duplexes. Consequently a 40 ps equilibration was performed with position constraints on the DNA duplexes. The production run was performed by using an NPT ensemble with periodic boundary conditions. The terminal bases of the DNA duplex were constrained using the SHAKE algorithm during the entire simulation. The production simulations were extended to 1 000 ps for all structures. RMSF values were analyzed using the 'Analyze Trajectory' function in Discovery Studio v19.1.0.18287. Final graphs were obtained with Origin Pro8 v8.0721 (OriginLab Corp., Northampton, MA, USA) (80).

### Circular dichroism spectroscopy

CD spectroscopy experiments were performed on a Chirascan V100 spectrometer using a quartz cuvette with an optical path length of 0.1 cm at 20°C. Spectra were collected between 340 and 200 nm, with a 1s sampling between time points. A blank sample containing only the buffer was treated in the same way and the spectra value was subtracted from the collected data. Oligonucleotide solutions of unliganded duplexes were prepared at a final concentration of 25 µM in buffer containing 20 mM MOPS (pH 6.5), 1 mM Na<sub>2</sub>-EDTA, and 50 mM NaCl. The samples were annealed by heating at 95°C for 10 min and cooled on ice for 30 min. To prepare the liganded DNA:ECHI complexes, 25 µM of duplexes were pre-incubated with 50 µM ECHI at 4°C for 24 h.

### NMR experiments

All NMR experiments were carried out on a BRUKER AVANCE 800 spectrometer equipped with a 5-mm triple resonance cryoprobe and a Z-axis pulsed field gradient at 5°C. For one-dimensional <sup>1</sup>H spectra, DNA duplexes samples were prepared at 1 mM in 20 mM NaCl, 20 mM sodium phosphate buffer (pH 7.0) with 10% D<sub>2</sub>O. NOESY experiments, with mixing times of 200 ms, were performed for the same samples but lyophilized and dissolved in 100% D<sub>2</sub>O. All NMR spectra were processed using Bruker Topspin software.

## Results

### Cytosine methylation at non-CpG sites induce polymorphic base pair geometry

The crystal structures of unmethylated d(ACGCCGT/ACGGCGT) (C:G pairing structure) and methylated d(ACG<sub>m</sub>CCGT/ACGGCGT) (<sub>m</sub>C:G pairing structure) were solved at resolutions of 2.35 and 2.0 Å, respectively. The data collection statistics are listed in [Supplementary Table S1](#). In these structures, oligonucleotides

self-assembled into antiparallel duplexes with four independent duplexes within each asymmetric unit (Figure 1A). The above duplexes were labelled C:G-NPX1 to C:G-NPX4 and  $m$ C:G-NPX1 to  $m$ C:G-NPX4 for unmethylated C:G and methylated  $m$ C:G pairing structures respectively. All duplexes of both the unmethylated and methylated structures were virtually superimposable, with an all-atom root-mean-square deviation (RMSD) of  $< 1 \text{ \AA}$  (Figure 1B). The central 5-bp segment of each duplex in both the unmethylated and methylated crystal structures have B-DNA-like duplex conformations. Consistent with the crystal structures, the circular dichroism (CD) spectra of both unmethylated and methylated structures showed negative and positive peaks at 250 and 275 nm, respectively, which are typical for B-DNA (Supplementary Figure S1).

Despite striking similarities among the unmethylated and methylated structures, we observed several differences between the methylated  $m$ C4:G11 base pairs and the unmethylated C4:G11 base pairs. To better describe these differences, common parameters including hydrogen bonding patterns, base 'pivot' angles ( $\lambda Y$  for pyrimidines and  $\lambda R$  for purines), C1'-C1' distances, and shearing distances were measured for the  $m$ C4:G11 and C4:G11 base pairs of each duplex and are listed in Supplementary Table S2 and described in Supplementary Note 1. The measured values suggest that the C4:G11 base pairs in all four unmethylated duplexes adopt a canonical Watson-Crick (WC) conformation (Figure 2A, Supplementary Figures S2A and S11A). On the other hand, six different conformations were observed for  $m$ C:G base pairs in the methylated duplexes: In  $m$ C:G-NPX1, the  $m$ C4:G11 base pair adopts two alternative conformations with occupancy ratios of 0.6 and 0.4, which were designated  $m$ C:G-NPX1-1 and  $m$ C:G-NPX1-2, respectively. In  $m$ C:G-NPX2,  $m$ C4:G11 base pair also refined to two alternative conformations with occupancy ratios of 0.9 and 0.1, which were thereby designated  $m$ C:G-NPX2-1 and  $m$ C:G-NPX2-2, respectively. The central  $m$ C4:G11 base pairs of  $m$ C:G-NPX3 and  $m$ C:G-NPX4 each adopt a WC conformation similar to the unmethylated structure (Figure 2B, Supplementary Figures S2B and S11B).

We further investigated the alternative conformations of the central  $m$ C4:G11 base pairs in  $m$ C:G-NPX1 and  $m$ C:G-NPX2. For  $m$ C:G-NPX1, both  $m$ C:G-NPX1-1 and  $m$ C:G-NPX1-2 adopt an asymmetric Watson-Crick (aWC) conformation (Figure 2C, Supplementary Figures S2C and S11C). Interestingly, for  $m$ C:G-NPX2, whilst  $m$ C4:G11 of  $m$ C:G-NPX2-1 adopts an aWC conformation similar to those observed for  $m$ C:G-NPX1-1 and  $m$ C:G-NPX1-2, the same base pair in  $m$ C:G-NPX2-2 adopts *anti-syn* pairing, in which N4 of  $m$ C4 forms a hydrogen bond with O6 of G11. Moreover, a water-mediated hydrogen bond is apparent between O2 of  $m$ C4 and N7 of G11 (Figure 2D, Supplementary Figures S2D and S11D). In addition, a bifurcated hydrogen bond was formed between O6 of G11 and protonated N3 of  $m$ C4 to further stabilize this *anti-syn* pairing. We have designated the unique geometry observed in  $m$ C:G-NPX2-2 as (w)cWH to represent a water-mediated *cis* Watson-Crick/Hoogsteen base pair geometry (81,82) (Supplementary Note 1). In summary, the base pair geometries of the central  $m$ C4:G11 in the methylated duplexes fall into three types: WC ( $m$ C:G-NPX3 and  $m$ C:G-NPX4), aWC ( $m$ C:G-NPX1-1,  $m$ C:G-NPX1-2, and  $m$ C:G-NPX2-1), and (w)cWH ( $m$ C:G-NPX2-2).

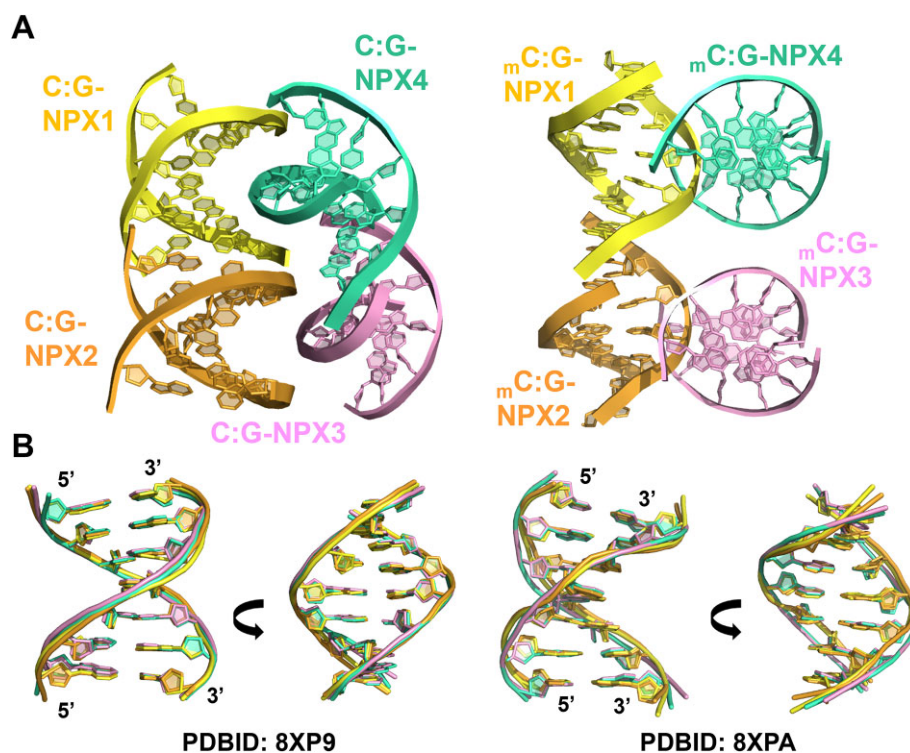
## Polymorphic base pair geometry resulted from non-CpG methylation rather than solely from WC base pair instability

The methylation-site polymorphisms suggested that cytosine methylation may enhance the dynamics of modified base pairs. In light of this, we performed MD simulations on the C:G-NPX1,  $m$ C:G-NPX3,  $m$ C:G-NPX2-1 and  $m$ C:G-NPX2-2 duplexes to investigate the impact of cytosine methylation on base flexibility, and calculated the root-mean-square fluctuation (RMSF) values in each structure to represent the flexibility of individual bases. The results showed that the largest fluctuation occurred at the central base of the methylated structure compared to the unmethylated structure (Supplementary Figure S3A). This indicates that cytosine methylation enhanced the base pair dynamics of methylated sites compared to the unmethylated structure, consistent with the high degree of polymorphism in methylated base pairs observed in the crystal structure.

To further test whether base pair dynamics are primarily responsible for the polymorphic base pair geometry that promote (w)cWH formation, we replaced the guanine of the center  $m$ C:G pair with inosine. Since the only difference between guanine and inosine is the absence of the N2 amino group in inosine, this substitution is ideal for reflecting the base pair instability associated with the loss of a single H-bond (83). The unliganded C:I pairing structure was determined at a resolution of 2.5 Å. Each asymmetric unit contains four independent duplexes, labelled C:I-NPX1 to C:I-NPX4. The central C4:I11 base pairs in all four duplexes adopt a single *anti-anti* pair in a non-canonical WC geometry (Supplementary Note 1 and Supplementary Figures S3B and S11E). This could be attributed to the instability of WC pairing due to the loss of H-bonds. However, we did not observe any alternative conformations in C:I-NPXs. To investigate the impact of inosine substitution on base flexibility, we performed MD simulation on the C:I-NPX1 structure. The choice of C:I-NPX1 as the representative structure was because its central C4:I11 base pair has parameter values that are close to the average values of the four duplexes. Our results showed that higher fluctuation was observed at the central bases of the C:I-NPX1 compared to the native unmethylated structure (Supplementary Figure S3A). This indicates that substitution of guanine by inosine enhanced base pair dynamics. Overall, the above results indicate that the instability of WC pairing could lead to higher base pair dynamics, while cytosine non-CpG methylation is mainly responsible for the observed polymorphic base pair geometry and (w)cWH formation.

## Structural basis of methylation-induced polymorphism at the $m$ C4:G11 position

As shown above, cytosine methylation at the non-CpG step may result in geometric polymorphism of methylated base pairs. The structural basis of this polymorphism may be deduced through a detailed analysis of the interaction differences among the unmethylated and methylated conformations. Because the geometry of the C4:G11 base pair in all four unmethylated duplexes were similar, C:G-NPX1 was selected as a representative of the unmethylated conformation. Likewise,  $m$ C:G-NPX3,  $m$ C:G-NPX2-1 and  $m$ C:G-NPX2-2 were chosen as representatives of duplexes containing WC, aWC, or (w)cWH base pairs, respectively. Figure 3A-C shows the



**Figure 1.** Crystal structures of the unliganded C:G and  $mC$ :G pairing duplexes. **(A)** Refined structures of unliganded C:G (left) and  $mC$ :G (right) pairing duplexes. Each asymmetric unit contains four independent duplexes. DNA duplexes are represented in cartoon form. **(B)** Superimposition of the crystal structures between C:G-NPX1/ $mC$ :G-NPX1 (yellow), C:G-NPX2/ $mC$ :G-NPX2 (orange), C:G-NPX3/ $mC$ :G-NPX3 (pink), and C:G-NPX4/ $mC$ :G-NPX4 (green) of native (left) and methylated (right) structures. The all-atom root mean square deviations (RMSDs) between C:G-NPX1 and C:G-NPX2, C:G-NPX3 and C:G-NPX4 are 0.2, 0.4, and 0.4, respectively. Meanwhile, the values between  $mC$ :G-NPX1 and  $mC$ :G-NPX2,  $mC$ :G-NPX3 and  $mC$ :G-NPX4 are 0.4, 0.9, and 0.8, respectively.

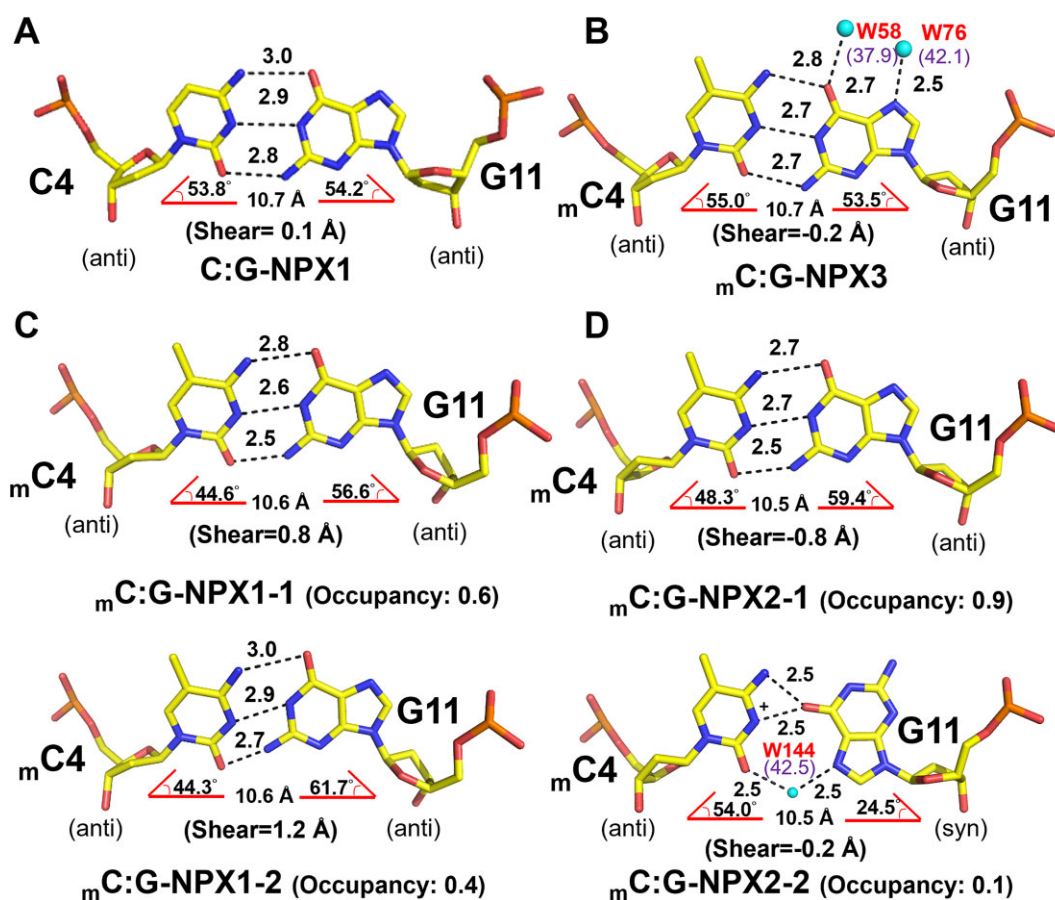
methyl- $\pi$  stacking interactions between the methyl groups of  $mC$  and adjacent nucleotides in the different methylated structures; interactions which were not available to the unmethylated DNA. The  $mC4$  and  $C5$  bases were better aligned when compared to their unmethylated counterparts. The  $\pi$ - $\pi$  distances between the  $mC4$  and  $C5$  bases in the methylated structures (3.5 Å) were also shorter than that in the unmethylated structure (3.9 Å) as shown in Figure 3D–G. These results suggest stronger stacking interactions in the methylated structures compared to the unmethylated structure. On the other hand, the distances between the fifth carbon of  $C4$  and the sugar ring of the  $G3$  base in the methylated structures were longer than those in the unmethylated counterparts (Supplementary Figure S4A–D). This difference is likely attributed to steric hindrance between the methyl group and the phosphate backbone, causing the methylated cytosine to move towards the interior of the helix. The repositioning of  $mC4$  sterically hinders WC geometry, making the methylated WC more flexible and facilitating the formation of (w)cWH geometry (Figure 3H).

The different methylated structures were further stabilized through distinct networks of metal ions and water molecules around their methylated base pairs (Supplementary Figure S4E–F). In  $mC$ :G-NPX3,  $G3$  and  $mC4$  are paired indirectly to each other through the W75- and W127-mediated interactions (Supplementary Figure S4E). A  $Mn^{2+}$  coordinates with two symmetry-related N7 atoms of  $G10$  and four water molecules W88, W57, W128, and W58 in a perfect octahedral geometry. This cluster of  $Mn^{2+}$  ion and water molecules stabilize the methylated base pairs of  $mC$ :G-NPX3 in a symmet-

rical WC conformation. In contrast, only one water molecule (W141) interacts with N7 of  $G3$  and N4 of  $mC4$  in both  $mC$ :G-NPX2-1 and  $mC$ :G-NPX2-2 (Supplementary Figure S4F). The smaller number of water-mediated interactions may provide more flexibility to the methylated base pair in  $mC$ :G-NPX2 compared to that in  $mC$ :G-NPX3, which may explain why the methylated base pair in  $mC$ :G-NPX2 could adopt two alternative conformations. In addition, the lack of water-mediated interactions stabilizing  $G10$  in  $mC$ :G-NPX2-1 may partially explain why it assumes an aWC conformation. In addition, we further observed that the stacking interaction between  $G11$  and  $G10$  seems to play a role in determining the conformation of  $G11$ . In the  $mC$ :G-NPX2-1 structure,  $G11$  is in the *anti*-form and aligned with  $G10$  in the same direction, whereas in the  $mC$ :G-NPX2-2 structure,  $G11$  is in the *syn*-form and is aligned with  $G10$  in the opposite direction, which may result in different stacking energies (Supplementary Figure S5). This could account for the difference in occupancy of the aWC and (w)cWH geometries in the  $mC$ :G-NPX2 duplex.

#### (w)cWH base pair formation can be stabilized by echinomycin intercalation

Since structural analysis indicated that stacking interaction between the guanine of the methylated  $mC$ :G base pair and the flanking bases plays an important role in stabilizing the guanine, we tried to further stabilize the methylated  $mC$ :G base pair using the drug echinomycin (ECHI). A single ECHI molecule intercalates into both sides of a 5'-CpG DNA step and forms strong stacking interactions with bases flanking the

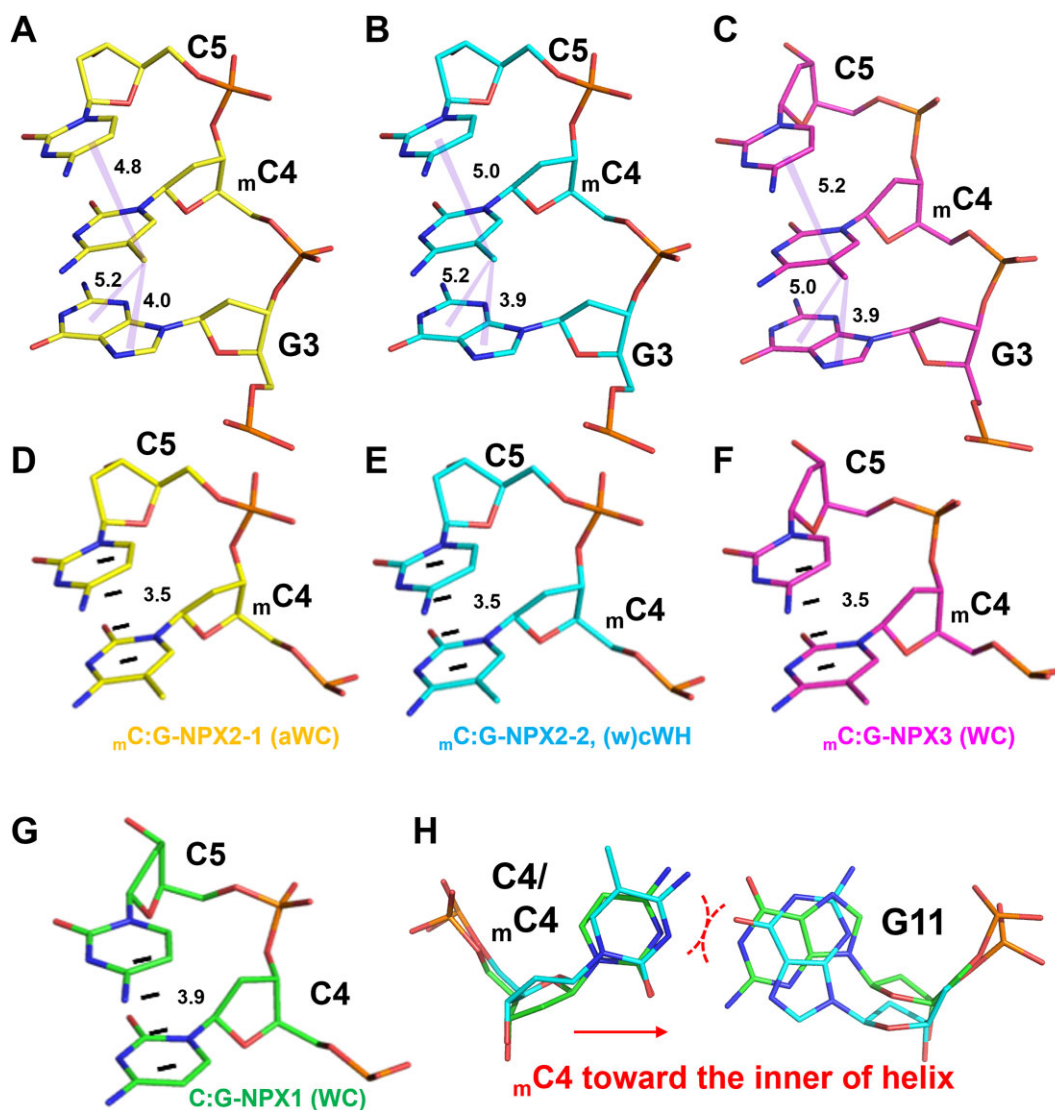


**Figure 2.** Geometries of the central base pairs in  $d(\text{ACGCCGT}/\text{ACGGCGT})_2$  with or without non-CpG methylated cytosine. Stick representation of the central C4:G11 base pairs of C:G-NPX1 (A) and  $m\text{C}4$ :G11 base pair of  $m\text{C}$ :G-NPX3 (B),  $m\text{C}$ :G-NPX1 (C) and  $m\text{C}$ :G-NPX2 (D). Water molecules appear as cyan spheres. Hydrogen bonds are represented by dotted lines, with numbers indicating the distance between the two contributing atoms in angstroms (Å). DNA parameters including C1'–C1' distance,  $\lambda$  angles, shear and occupancy are indicated at the bottom of the figure. B-factors of each water molecule are indicated with purple letters in brackets.

intercalation site (84) (Figure 4A). The crystal structures of C:G and  $m\text{C}$ :G duplexes in complex with ECHI were designated C:G-EPX and  $m\text{C}$ :G-EPX, and solved at 2.0 and 1.64 Å resolution, respectively (Supplementary Table S1). Overall, the two structures are closely similar, with an all-atom RMSD of 0.18 Å, and their conformations resemble previously proposed DNA:ECHI complex structures (74) (Figure 4B). The CD spectra of C:G-EPX and  $m\text{C}$ :G-EPX structures revealed that these structures are nearly identical in solution (Supplementary Figure S6), on accord with the results of the crystal structure analysis. Each asymmetric unit contains a DNA duplex bound by two ECHI molecules, labeled ECHI 1 and ECHI 2. The two planar quinoxaline rings of each ECHI molecule flank the CpG steps in the minor groove (Figure 4B). The details of ECHI binding to C:G and  $m\text{C}$ :G duplexes are consistent with previously reported echinomycin-DNA interactions (Figure 4C and Supplementary Note 2).

The geometries of the  $m\text{C}4$ :G11 and C4:G11 base pairs in each complex are fully described in Supplementary Note 2, with the parameters listed in Supplementary Table S2. The C4:G11 base pair in the C:G-EPX structure adopts a symmetrical WC conformation similar to that of the C:G-NPX1 structure (Figure 5A, Supplementary Figures S7A and S12A). On the other hand, the  $m\text{C}4$ :G11 base pair in the  $m\text{C}$ :G-EPX structure exhibits two alternative conformations with occupancies

of 0.7 and 0.3, designated  $m\text{C}$ :G-EPX-1 and  $m\text{C}$ :G-EPX-2, respectively. The  $m\text{C}4$ :G11 base pair in  $m\text{C}$ :G-EPX-1 adopts a symmetrical WC conformation similar to the unmethylated structure (Figure 5B, Supplementary Figures S7B and S12B). The central base pair of the  $m\text{C}$ :G-EPX-2 structure adopts a geometry similar to (w)cWH (Figure 5C, Supplementary Figures S7C and S12B). The *syn*-G11 in  $m\text{C}$ :G-EPX-2 was found to stack better with the QUI ring of ECHI2 compared to *anti*-G11 in C:G-EPX and  $m\text{C}$ :G-EPX-1 (Supplementary Figure S8A–C). In addition, water molecules in the vicinity of  $m\text{C}4$ :G11 are organized differently in C:G-EPX,  $m\text{C}$ :G-EPX-1 and  $m\text{C}$ :G-EPX-2 (Supplementary Figure S8D–F). A single conserved water molecule is present in both C:G-EPX and  $m\text{C}$ :G-EPX complexes (W8 in C:G-EPX and W6' in both  $m\text{C}$ :G-EPX-1 and  $m\text{C}$ :G-EPX-2). This water molecule was found to mediate the indirect interactions between N4 of the ECHI1 quinoline ring and O6 of G11 in both the C:G-EPX and  $m\text{C}$ :G-EPX complexes. In the latter complexes, W6' also bridges N4 of the ECHI1 quinoline and N4 of  $m\text{C}4$ , further stabilizing the position of the  $m\text{C}4$  base. Three additional water molecules, W1, W6 and W41, were observed only in the methylated complexes. W41 is the only one common to both  $m\text{C}$ :G-EPX-1 and  $m\text{C}$ :G-EPX-2, and mediates the interaction between N4 of the ECHI2 quinoline and O6 of G11 (Supplementary Figure S8B, C). W1 and W6 are unique to



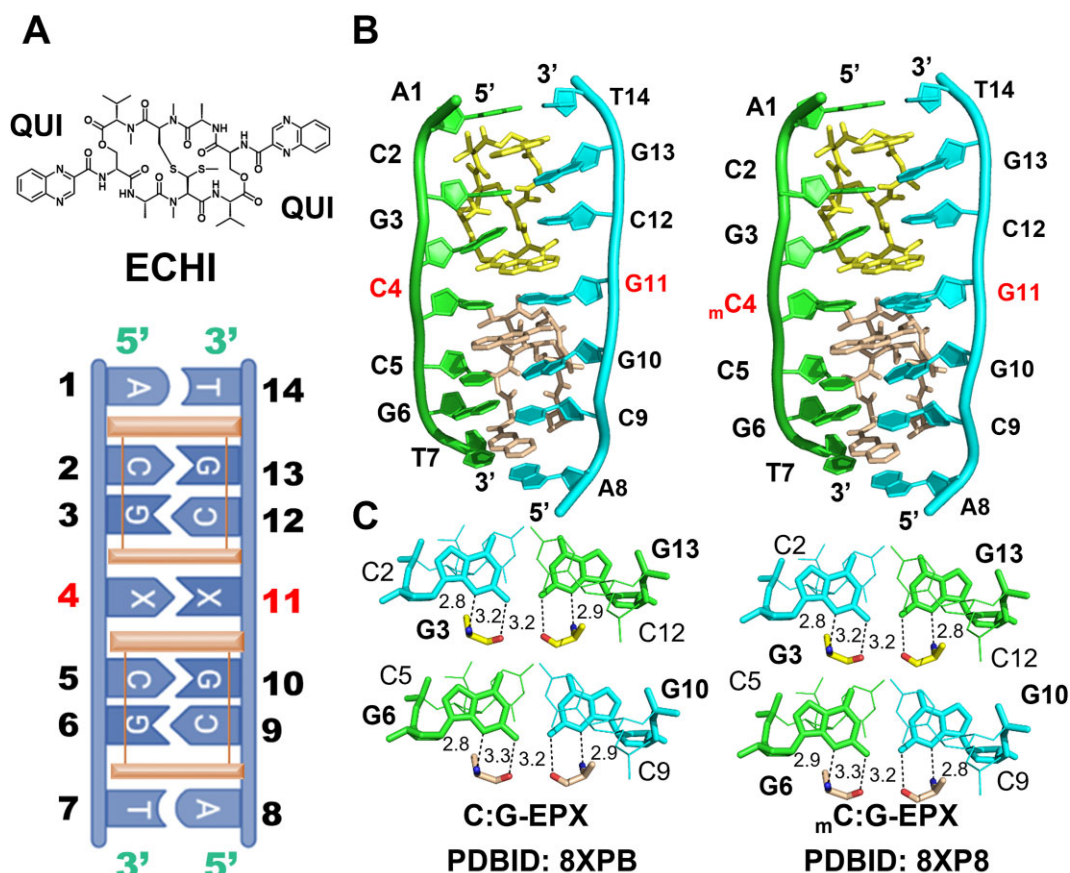
**Figure 3.** Effect of stacking interactions involving 5<sub>m</sub>C on the geometry of DNA base pairs. (A–C) Stacking of methyl groups with G3 and C5 bases in *m*C:G-NPX2-1 (A), *m*C:G-NPX2-2 (B) and *m*C:G-NPX3 (C) structures. (D–G)  $\pi$ - $\pi$  stacking between *m*C4/C4 and C5s in C:G-NPX1 (D), *m*C:G-NPX3 (E), *m*C:G-NPX2-1 (F) and *m*C:G-NPX2-2 (G) are indicated by the black dashed lines. (H) Changes of *m*C4 position sterically destabilize *m*C:G WC bps (red dashes) and favor the (w)cWH base pair formation.

*m*C:G-EPX-2. W6 interacts with W41 and N1 of G11, helping to stabilize G11 in the *syn*-conformation. W1 interacts with O2 in *m*C4 and N7 in G11, further helping to stabilize the unusual geometry of the (w)cWH base pair in *m*C:G-EPX-2. A hydrogen bond was also found between N2 and OP2 of G11, which stabilizes the *syn*-conformation of G11 in *m*C:G-EPX-2 (Supplementary Figure S8C). Collectively, these extra interactions may play a significant role in stabilizing the *m*C4:G11 base pair in *m*C:G-EPX-1 and *m*C:G-EPX-2, while the unique hydrogen bonds observed in *m*C:G-EPX-2 may account for the increased occupancy of the (w)cWH geometry observed in ECHI-bound DNA.

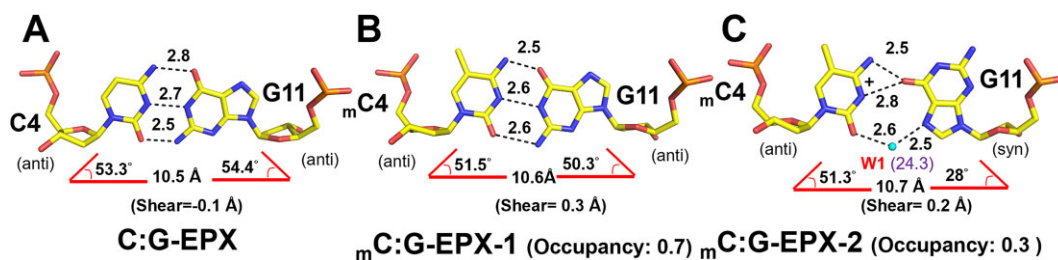
#### Alternative conformations are observed for non-CpG methylated duplexes in solution

The non-CpG methylation-induced base pair dynamics and (w)cWH formation were monitored by NMR experiments. To avoid the fraying effect at the terminal bases that could im-

pair the DNA double helix conformation during NMR experiments, we incorporated two additional base pairs flanking the *m*C:G/C:G pairing duplexes to enhance their stability (Figure 6A). The duplexes, d(GAACGCCGTAC/GTACGGCGTTC)<sub>2</sub> and d(GAACG<sub>m</sub>CCGTAC/GTACGGCGTTC)<sub>2</sub>, were designated LC:G and L<sub>m</sub>C:G, respectively. Figure 6B shows the 1D <sup>1</sup>H NMR spectra of L<sub>m</sub>C:G and LC:G structures. From the spectra, it is clear that the unmethylated LC:G structure most likely forms a duplex, confirmed by the observation of six sharp signals in the region between 12 and 13 ppm, as well as four signals within the 13 to 14 ppm range. These signals are characteristic of the H1 proton of a guanine residue and the H3 proton of thymine involved in WC base pairs, respectively. The spectrum of the methylated L<sub>m</sub>C:G structure exhibited a similar signal pattern in the imino region, except for the signals observed between 12.65 and 12.75 ppm. The signals of the G16 and G17 imino protons, which occur at 12.71 and 12.65 ppm in the LC:G spectrum, shifted and overlapped into one signal at 12.73 ppm in the L<sub>m</sub>C:G spectrum



**Figure 4.** Crystal structures of DNA:ECHI complex with or without methylated cytosine. **(A)** (upper) Chemical structure of ECHI. (lower) Schematic representation of the DNA:ECHI complex. Central X4:X11 (highlighted in bold red) represent different central base pairs used in this study. ECHI intercalated between the C2pG3/C5pG6 steps is represented in orange. **(B)** Overall structures of C:G-EPX (left) and  $m$ C:G-EPX (Right) complex. ECHI1 and ECHI2 are colored in yellow and orange, respectively. **(C)** Hydrogen bonds between ECHI and DNA at the CpG step of C:G-EPX (left) and  $m$ C:G-EPX (Right) complexes. Dotted lines represent direct interactions between atoms, and numbers indicate the distances.



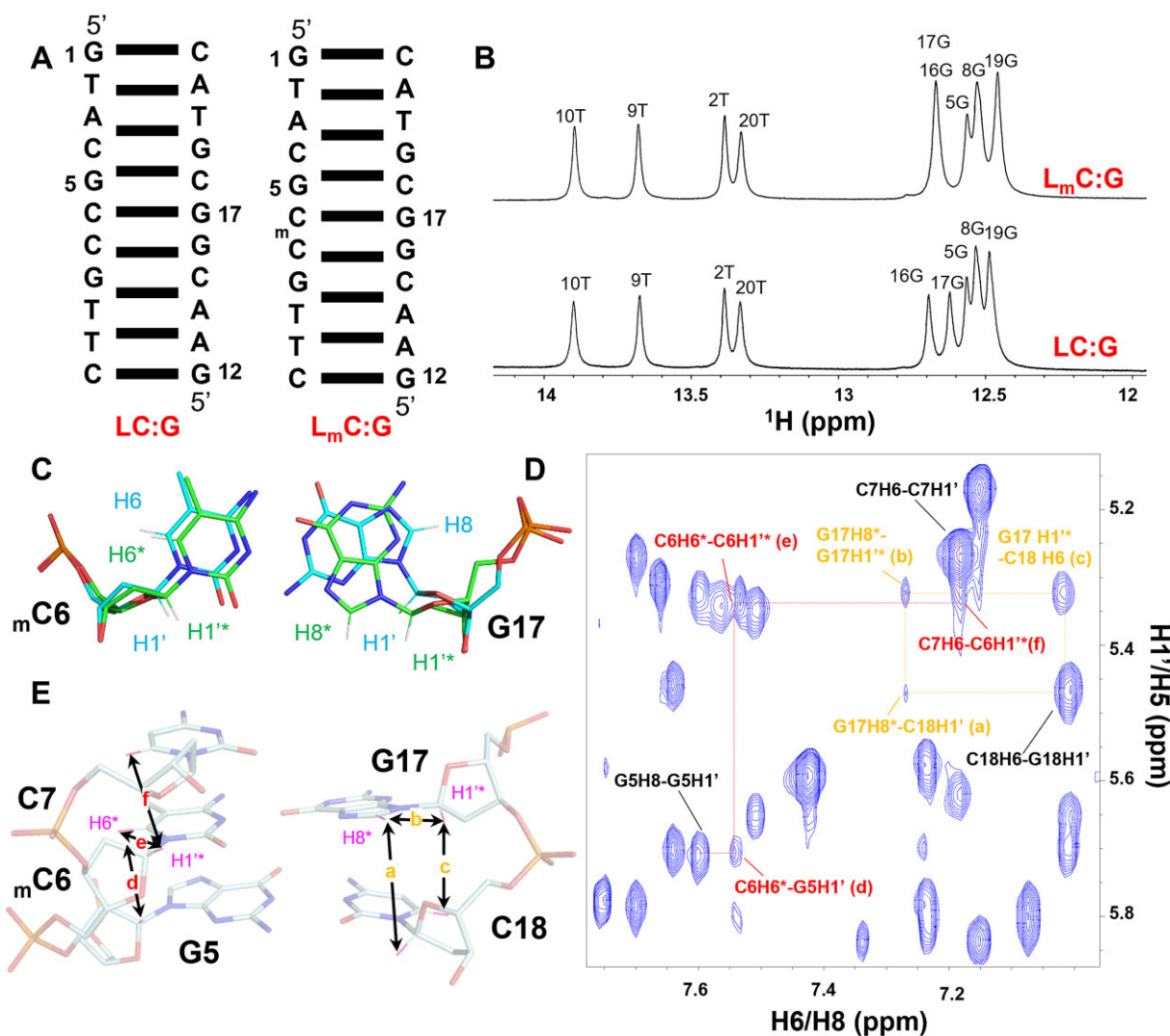
**Figure 5.** Central base pair geometry in C:G-EPX and  $m$ C:G-EPX complexes. **(A–C)** Stick representation of the central base pairs of C:G-EPX **(A)**,  $m$ C:G-EPX-1 **(B)** and  $m$ C:G-EPX-2 **(C)**. Water molecules appear as cyan spheres. Hydrogen bonds are represented by dotted lines, with numbers indicating the distance between two contributing atoms in angstroms (Å). DNA parameters including C1'-C1' distance,  $\lambda$  angles, shear and occupancy are indicated at the bottom of the figure. B-factors of each water molecule are indicated with purple letters in brackets.

(Figure 6B). This observation suggests that the overall folds of both duplexes were similar with the exception of G16 and G17, which were affected by the presence of the methyl group in  $m$ C.

To further explore the conformational differences induced by non-CpG methylation, two-dimensional nuclear Overhauser effect spectroscopy (NOESY) experiments were performed. **Supplementary Figures S9A** and **B** shows the aromatic to H1'/H5 region of the NOESY in D<sub>2</sub>O for the LC:G and L<sub>m</sub>C:G structures, respectively, which were assigned using well-established strategies (85). A comparison of the NOE patterns between LC:G and L<sub>m</sub>C:G within this region is

shown in **Supplementary Figure S9C**. Four additional NOEs (peaks a-b and d-e), none of which were part of sequential walks, were observed in the L<sub>m</sub>C:G spectrum, indicating that an additional conformation might be present in the methylated duplex. We inferred that these additional signals may belong to the alternative (w)cWH conformation of the methylated base-pair. This inference was confirmed by the six NOEs, correlated to the H6 and H1' protons in cytosine, as well as H8 and H1' protons in guanine of the alternative (w)cWH conformation of methylated base-pair (Figure 6C). These include the four unique NOEs observed only in L<sub>m</sub>C:G spectrum: between the G17H8\* proton and the



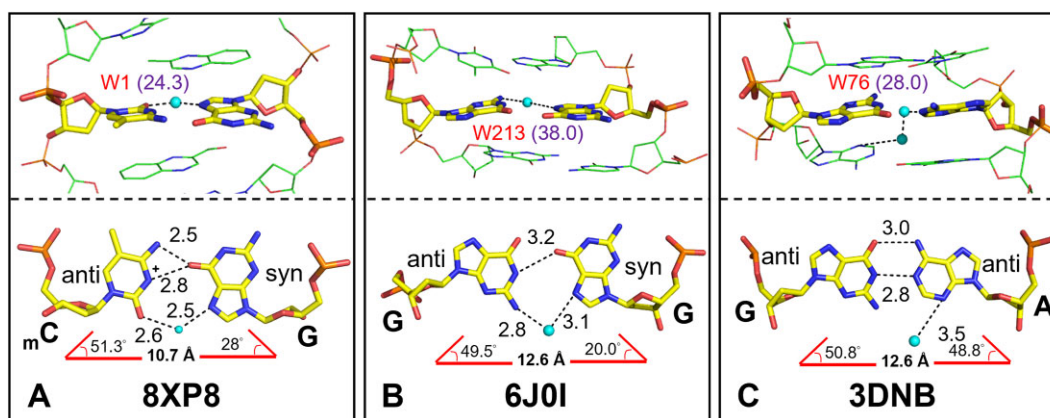


**Figure 6.** non-CpG methylation-induced alternative conformations in solution. **(A)** LC:G and L<sub>m</sub>C:G DNA duplexes used in this study. **(B)** 1D imino <sup>1</sup>H spectra of LC:G (lower) and L<sub>m</sub>C:G (upper). Spectra were measured in 20 mM NaCl, 20 mM PBS buffer (pH = 7) at 5°C, and 1 mM DNA duplex concentrations. **(C)** Schematic illustration of the H8, H6, and H1' protons in major (blue) and alternative (w)cWH (green) conformations of the methylated base-pair. **(D)** 2D <sup>1</sup>H, <sup>1</sup>H NOESY spectra of L<sub>m</sub>C:G. Six crucial NOEs characteristic of an alternative (w)cWH conformation of mC6:G17 base-pair, as indicated in **(E)** were marked with lowercase letters a–f in brackets.

C18H1' proton (cross-peak a), the G17H8\* proton and the G17H1\*\* proton (cross-peak b, which was overlapped with the NOE between A3H2 and C4H1'), the C6H6\* proton and the G5H1' proton (cross-peak d) and between the C6H6\* proton and the C6H1\*\* proton (cross-peak e). As well as the NOE exhibited from C7H6 proton to the C6H1\*\* proton (cross-peak f), and which exhibited from G17H1\*\* proton to C18H6 proton (cross-peak c, overlapped with the NOE exhibited from C4H8 to C4H1') (Figure 6D–E). In the alternative (w)cWH conformation of the methylated base-pair, it was observed that G17 adopted a *syn*-conformation. This was supported by the G17H8\*–C18H1' connectivity (peak a), which is typical for *syn-anti* step of a sequential walk (Figure 6E). The occurrence of an alternative conformation suggested non-CpG methylation enhanced dynamics of the methylated base-pair is consistent with our crystallography results. In total, the NMR data provide a hint of the existence of an *anti*→*syn* transition of guanosine in the non-CpG methylated mC:G base pair in solution, which may indicate (w)cWH formation.

## Discussion

Several previous studies have shown that cytosine methylation affects base flexibility, which in turn impacts DNA local structure (86–88). To investigate the structural impact of non-CpG methylation on the structure of duplex DNA, we have determined the crystal structures of DNA sequences associated with neurodevelopmental disorder with or without a non-CpG methylation site. The crystal structure of the methylated mC:G duplex has revealed that cytosine methylation at non-CpG site results in geometric polymorphism of the methylated mC:G base pair, which leads to the formation of a unique (w)cWH geometry with a *syn*-guanosine in a population of around 10%. This phenomenon was also confirmed by NMR spectroscopy in aqueous solution. Molecular dynamics (MD) simulations of the unliganded structures further demonstrate that non-CpG methylation may enhance the dynamics of methylated base pairs. These findings imply that in a non-CpG context, the methylated cytosine is highly dynamic which in turn may allow the *anti*→*syn* transition of its



**Figure 7.** The role of water molecules in stabilizing non-canonical base pairs. Water-mediated interactions stabilize unusual base pair geometries by avoiding steric clashes between bases. Non-canonical base pairs are shown in yellow sticks. Water molecules appear as a cyan spheres. Hydrogen bonds are represented by dotted lines, with numbers indicating the distance between two contributing atoms in angstroms (Å). The C1'–C1' distance and  $\lambda$  angles are indicated at the bottom of the figure. The respective PDB-ID for each structure is given in bold black letters. *B*-factors of each water molecule are indicated with purple letters in brackets.

complementary guanosines, resulting in the formation of the (w)cWH geometry.

The source of the highly dynamic character of the methylated cytosine may originate from several causes. The identity of the 3'-neighboring base next to the methylation site may play an important role because a bulkier base (e.g. G or A) may constrain the dynamic freedom of the methylated cytosine. The transition from WC to (w)cWH also requires a large space, which may be defined by the intra-strand phosphate-to-phosphate (P-P) distances of adjacent nucleotides. B-DNA has the largest P-P distance of all forms of DNA, and may be a prerequisite for the conformational transition to occur (66). On the other hand, base dynamics of the methylated cytosine may not be sufficient to form the (w)cWH geometry. Using the unmethylated sequence, we introduced additional base pair instability by substituting the guanosine with an inosine (83), which results in the loss of an H-bond, leading to only a single base pair geometry in the structure, even though the dynamics of the C:I base pair had been enhanced compared to the original sequence. This suggests that either the base dynamics were not enhanced to a sufficient degree to allow the (w)cWH geometry to occur, or that base dynamics alone may not be sufficient to cause the (w)cWH geometry to form.

Further clues lie in the atomic interactions between the bases surrounding the methylation site. Our structural analyses showed that in the non-CpG step, the stacking interactions between  $m$ C:G base pair and its adjacent nucleotides are crucial for stabilizing the (w)cWH configuration. For example, the bulky base at the G10 position of our DNA sequence constrains G11 to be in the *syn*-conformation. This is corroborated by the observation that the echinomycin-bound structures, which have enhanced stacking interactions caused by ligand intercalation, have higher (w)cWH base-pair occupancies compared to the unliganded structures. Taken together, stabilization of the (w)cWH geometry appears to employ all the hallmark mechanisms found in the stabilization of a typical Hoogsteen base pair.

Based on our findings, an analysis of all available structures of unliganded DNA duplexes containing methylated cytosines may provide clues on why the (w)cWH geometry

has never been previously reported (Supplementary Table S3) (36,38,39,73,89–92). Among all DNA sequences studied, that used in this study is the only one that fulfilled all of the following criteria: (a) having the dynamic freedom of the methylated cytosine; (b) forms a B-form DNA; (c) the Hoogsteen base pair has favorable stacking interactions between the methylated base pair and adjacent bases. Accordingly, we propose that the (w)cWH geometry may be specific to base steps CpC and CpT, which fulfils all of the above criteria (Supplementary Figure S10). Our crystal structures provide the first insight of the alternative (w)cWH geometry arising from non-CpG methylation.

X-ray crystallography is a powerful tool for identifying alternative conformations and pinpointing the atomic positions of water molecules in biological macromolecules (93,94). Miao and Cao analysed protein structures containing alternative side-chain conformations and concluded that alternative conformation could be observed in crystal structures when the resolution is better than 2.0 Å (95), further rationalizing the discovery of alternative conformations in our high-resolution data. In the unliganded (w)cWH base pair, the electron density associated with one water molecule at 0.12 occupancy may be insufficient to confidently determine its precise position. However, with the increased occupancy of 0.32 upon stabilization of (w)cWH by ECHI, a highly similar geometry as well as water location accompanied by a clearer map was observed. This strengthens the credibility of the water molecule location within the (w)cWH in both the unliganded and liganded structures.

The importance of water cannot be overstated. Previous studies have shown that cytosine methylation affected the hydration pattern around the methylated sites, which may be important for protein recognition (96,97). In this work, although the *syn*-guanosine is engaged in Hoogsteen-edge pairing with methylated cytosine to reduce steric clashes between the bases, the longer C1'–C1' distance results in weaker hydrogen bonds compared to canonical HG base pairs. These weaker hydrogen bonds are compensated by water-mediated interactions (Figure 7A). Other studies have shown that two mismatches with steric clashes, G:G and G:A, can generate unusual base-pair geometries through water-mediated pairing, with the G:G

mismatch adopting a water stabilized *anti-syn* geometry (Figure 7B) (54), and the G:A mismatch adopting an unusual geometry with a large propeller twist mediated by two water molecules to avoid steric clashes between mismatched bases (Figure 7C) (98). These results imply that unusual base pairings that are stabilized through water-mediated interactions may be more prevalent than one might expect.

The ability of non-CpG methylated DNA to adopt the (w)cWH geometry may have important functional consequences in epigenetic regulation. Several cancers have been associated with methylation status changes at non-CpG sites of the genome (10,32,99). Although spontaneous deamination is usually invoked as the major mutagenic mechanism in these cases, our findings allude to the intriguing possibility of an alternative mutagenic mechanism: methylation at the non-CpG sites may promote formation of (w)cWH base pairs, which are less stable than WC base pairs and have been shown to increase the susceptibility of double-stranded DNA to damage (100,101). The (w)cWH base pairs may also affect gene regulation because Hoogsteen base pairs are known to play important roles in DNA-protein recognition (59,60,63,65).

In summary, we have identified a novel (w)cWH geometry that may form part of the conformational ensemble specific to non-CpG methylated DNA. We propose that this (w)cWH geometry may provide an alternative step in the mechanism for methylation-mediated genetic instability and may have important functional consequences in epigenetic regulation. This study thus provides the first structural insights into the formation of alternative base pair geometries in non-CpG methylation.

## Data availability

All data are available in the main text or the supplementary materials. Structures have been deposited in the PDB (<https://www.rcsb.org/>) under accession numbers 8XP9 (unliganded C:G pairing structure), 8XPA (unliganded mC:G pairing structure), 8WNB (unliganded C:I pairing structure), 8XPB (C:G-ECHI complex) and 8XP8 (mC:G-ECHI complex).

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

We acknowledge the National Synchrotron Radiation Research Center (Taiwan) staff for X-ray data collection. NMR spectra were obtained at the High-Field NMR Center supported by Academia Sinica Core Facility and Innovative Instrument Project (AS-CFII-111-214).

*Author contribution:* Conceptualization: Shan-Meng Lin, Hsiang-Ti Huang and Ming-Hon Hou. Methodology: Shan-Meng Lin, Hsiang-Ti Huang. Investigation: Shan-Meng Lin and Hsiang-Ti Huang, Pei-Ju Fang and Chi-Fon Chang. Visualization: Shan-Meng Lin and Ming-Hon Hou. Supervision: Stephen Neidle and Ming-Hon Hou. Writing-original draft: Shan-Meng Lin, Roshan Satange, Stephen Neidle and Ming-Hon Hou. Writing-review & editing: Shan-Meng Lin, Shan-Ho Chou, Chung-ke Chang, Stephen Neidle and Ming-Hon Hou

## Funding

Ministry of Science and Technology, Taiwan, R.O.C. [109-2628-M-005-001-MY4, 109-2311-B-005-007-MY3 to M.-H.H.]. Funding for open access charge: Ministry of Science and Technology, Taiwan.

## Conflict of interest statement

None declared.

## References

- Schübeler, D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.
- Moore, L.D., Le, T. and Fan, G. (2013) DNA methylation and its basic function. *Neuropsychopharmacology*, **38**, 23–38.
- Haghani, A., Li, C.Z., Robeck, T.R., Zhang, J., Lu, A.T., Ablaeva, J., Acosta-Rodríguez, V.A., Adams, D.M., Alagaili, A.N., Almunia, J., et al. (2023) DNA methylation networks underlying mammalian traits. *Science*, **381**, eabq5693.
- Kato, M., Miura, A., Bender, J., Jacobsen, S.E. and Kakutani, T. (2003) Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr. Biol.*, **13**, 421–426.
- Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
- Esteller, M. (2005) Aberrant DNA methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.*, **45**, 629–656.
- Kandi, V. and Vadakedath, S. (2015) Effect of DNA methylation in various diseases and the probable protective role of nutrition: a mini-review. *Cureus*, **7**, e309.
- Krolevets, M., Cate, V.t., Prochaska, J.H., Schulz, A., Rapp, S., Tenzer, S., Andrade-Navarro, M.A., Horvath, S., Niehrs, C. and Wild, P.S. (2023) DNA methylation and cardiovascular disease in humans: a systematic review and database of known CpG methylation sites. *Clin. Epigenetics*, **15**, 56.
- Lanata, C.M., Chung, S.A. and Criswell, L.A. (2018) DNA methylation 101: what is important to know about DNA methylation and its role in SLE risk and disease heterogeneity. *Lupus Sci. Med.*, **5**, e000285.
- Ramasamy, D., Deva Magendhra Rao, A.K., Rajkumar, T. and Mani, S. (2021) Non-CpG methylation—a key epigenetic modification in cancer. *Brief. Funct. Genomics*, **20**, 304–311.
- Srivastava, R. and Lodhi, N. (2022) DNA methylation malleability and dysregulation in cancer progression: understanding the role of PARP1. *Biomolecules*, **12**, 417.
- Truong, M., Yang, B., Wagner, J., Desotelle, J. and Jarrard, D.F. (2013) Analysis of promoter non-CG methylation in prostate cancer. *Epigenomics*, **5**, 65–71.
- Nishiyama, A. and Nakanishi, M. (2021) Navigating the DNA methylation landscape of cancer. *Trends Genet.*, **37**, 1012–1027.
- Das, P.M. and Singal, R. (2004) DNA methylation and cancer. *J. Clin. Oncol.*, **22**, 4632–4642.
- Jin, Z. and Liu, Y. (2018) DNA methylation in human diseases. *Genes Dis.*, **5**, 1–8.
- Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
- Greenberg, M.V.C. and Bourc'his, D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.*, **20**, 590–607.
- Lyko, F. (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.*, **19**, 81–92.
- Jurkowska, R.Z., Jurkowski, T.P. and Jeltsch, A. (2011) Structure and function of mammalian DNA methyltransferases. *ChemBioChem*, **12**, 206–222.

20. Patil,V, Ward,R.L. and Hesson,L.B. (2014) The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics*, **9**, 823–828.
21. Ziller,M.J., Müller,F., Liao,J., Zhang,Y., Gu,H., Bock,C., Boyle,P., Epstein,C.B., Bernstein,B.E., Lengauer,T., *et al.* (2011) Genomic distribution and inter-sample variation of non-CpG methylation across Human cell types. *PLoS Genet.*, **7**, e1002389.
22. de Mendoza,A., Poppe,D., Buckberry,S., Pflueger,J., Albertin,C.B., Daish,T., Bertrand,S., de la Calle-Mustienes,E., Gómez-Skarmeta,J.L., Nery,J.R., *et al.* (2021) The emergence of the brain non-CpG methylation system in vertebrates. *Nat. Ecol. Evol.*, **5**, 369–378.
23. Lutz,P.-E., Chay,M.-A., Pacis,A., Chen,G.G., Aouabed,Z., Maffioletti,E., Théroux,J.-F., Grenier,J.-C., Yang,J., Aguirre,M., *et al.* (2021) Non-CG methylation and multiple histone profiles associate child abuse with immune and small GTPase dysregulation. *Nat. Commun.*, **12**, 1132.
24. Sereshki,S., Lee,N., Omirou,M., Fasoula,D. and Lonardi,S. (2023) On the prediction of non-CG DNA methylation using machine learning. *NAR Genom. Bioinform.*, **5**, lqad045.
25. Cao,Q., Wang,S., Ma,H., Luo,S., He,H. and Zhang,Y. (2023) Hyper non-CG methylation of expanded plant disease resistance NLR genes. *Plant Cell Rep.*, **42**, 1251–1254.
26. Wu,Z., Li,C., Zhu,R., Cao,Y., Chen,T.C. and Cheng,L. (2023) Reduced non-CpG methylation is a potential epigenetic target after spinal cord injury. *Neural Regener. Res.*, **18**, 2489–2496.
27. Nishino,K., Hattori,N., Sato,S., Arai,Y., Tanaka,S., Nagy,A. and Shiota,K. (2011) Non-CpG methylation occurs in the regulatory region of the sry gene. *J. Reprod. Dev.*, **57**, 586–593.
28. Jang,H.S., Shin,W.J., Lee,J.E. and Do,J.T. (2017) CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel)*, **8**, 148.
29. Ramsahoye,B.H., Biniszkiwicz,D., Lyko,F., Clark,V., Bird,A.P. and Jaenisch,R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Nat. Acad. Sci. U.S.A.*, **97**, 5237–5242.
30. Stroud,H., Do,T., Du,J., Zhong,X., Feng,S., Johnson,L., Patel,D.J. and Jacobsen,S.E. (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.*, **21**, 64–72.
31. Kinde,B., Gabel,H.W., Gilbert,C.S., Griffith,E.C. and Greenberg,M.E. (2015) Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6800–6806.
32. Xiao,W., Liu,X., Niu,X., Li,C., Guo,Y., Tan,J., Xiong,W., Fan,L. and Li,Y. (2020) The frequency of CpG and non-CpG methylation of Notch3 gene promoter determines its expression levels in breast cancer cells. *Exp. Cell Res.*, **386**, 111743.
33. Tillotson,R., Cholewa-Waclaw,J., Chhatbar,K., Connelly,J.C., Kirschner,S.A., Webb,S., Koerner,M.V., Selfridge,J., Kelly,D.A., De Sousa,D., *et al.* (2021) Neuronal non-CG methylation is an essential target for MeCP2 function. *Mol. Cell*, **81**, 1260–1275.
34. Amir,R.E., Van den Veyver,I.B., Wan,M., Tran,C.Q., Francke,U. and Zoghbi,H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185–188.
35. Chen,L., Chen,K., Lavery,L.A., Baker,S.A., Shaw,C.A., Li,W. and Zoghbi,H.Y. (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Nat. Acad. Sci. U.S.A.*, **112**, 5509–5514.
36. Fujii,S., Wang,A.H., van der Marel,G., van Boom,J.H. and Rich,A. (1982) Molecular structure of (m5 dC-dG)<sub>3</sub>: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.*, **10**, 7879–7892.
37. Genest,D., Hartmann,B., Thuong,N.T., Ptak,M. and Leng,M. (1984) Thermal stability of the Z-conformation of the tetranucleoside triphosphate (m5dC-dG)<sub>2</sub>. *Biochem. Biophys. Res. Commun.*, **125**, 803–811.
38. Vargason,J.M., Eichman,B.F. and Ho,P.S. (2000) The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nat. Struct. Biol.*, **7**, 758–761.
39. Mayer-Jung,C., Moras,D. and Timsit,Y. (1997) Effect of cytosine methylation on DNA-DNA recognition at CpG steps. *J. Mol. Biol.*, **270**, 328–335.
40. Robertson,K.D. (2002) DNA methylation and chromatin – unraveling the tangled web. *Oncogene*, **21**, 5361–5379.
41. Kiselev,I.S., Kulakova,O.G., Boyko,A.N. and Favorova,O.O. (2021) DNA methylation As an epigenetic mechanism in the development of multiple sclerosis. *Acta Naturae*, **13**, 45–57.
42. Cheng,X., Hashimoto,H., Horton,J.R. and Zhang,X. (2011) In: Tollefsbol,T. (ed.) *Handbook of Epigenetics*. Academic Press, San Diego, pp. 9–24.
43. Buitrago,D., Labrador,M., Arcon,J.P., Lema,R., Flores,O., Esteve-Codina,A., Blanc,J., Villegas,N., Bellido,D., Gut,M., *et al.* (2021) Impact of DNA methylation on 3D genome structure. *Nat. Commun.*, **12**, 3243.
44. Fang,J., Jiang,J., Leichter,S.M., Liu,J., Biswal,M., Khudaverdyan,N., Zhong,X. and Song,J. (2022) Mechanistic basis for maintenance of CHG DNA methylation in plants. *Nat. Commun.*, **13**, 3877.
45. Du,J., Johnson,L.M., Groth,M., Feng,S., Hale,C.J., Li,S., Vashisht,A.A., Wohlschlegel,J.A., Patel,D.J. and Jacobsen,S.E. (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol. Cell*, **55**, 495–504.
46. Li,X., Harris,C.J., Zhong,Z., Chen,W., Liu,R., Jia,B., Wang,Z., Li,S., Jacobsen,S.E. and Du,J. (2018) Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E8793–E8802.
47. Qiu,Q., Mei,H., Deng,X., He,K., Wu,B., Yao,Q., Zhang,J., Lu,F., Ma,J. and Cao,X. (2019) DNA methylation repels targeting of Arabidopsis REF6. *Nat. Commun.*, **10**, 2063.
48. Yang,J., Horton,J.R., Wang,D., Ren,R., Li,J., Sun,D., Huang,Y., Zhang,X., Blumenthal,R.M. and Cheng,X. (2019) Structural basis for effects of CpA modifications on C/ebpβ binding of DNA. *Nucleic Acids Res.*, **47**, 1774–1785.
49. Abhishek,S., Nakarakanti,N.K., Deeksha,W. and Rajakumara,E. (2021) Mechanistic insights into recognition of symmetric methylated cytosines in CpG and non-CpG DNA by UHRF1 SRA. *Int. J. Biol. Macromol.*, **170**, 514–522.
50. Fang,J., Leichter,S.M., Jiang,J., Biswal,M., Lu,J., Zhang,Z.-M., Ren,W., Zhai,J., Cui,Q., Zhong,X., *et al.* (2021) Substrate deformation regulates DRM2-mediated DNA methylation in plants. *Sci. Adv.*, **7**, eabd9224.
51. Olson,W.K., Bansal,M., Burley,S.K., Dickerson,R.E., Gerstein,M., Harvey,S.C., Heinemann,U., Lu,X.J., Neidle,S., Shakked,Z., *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.
52. Ruzkowska,A., Zheng,Y.Y., Mao,S., Ruzkowski,M. and Sheng,J. (2021) Structural insights into the 5'UG/3'GU wobble tandem in complex with Ba(2+) cation. *Front Mol. Biosci.*, **8**, 762786.
53. Bansal,A., Kaushik,S. and Kukreti,S. (2022) Non-canonical DNA structures: diversity and disease association. *Front. Genet.*, **13**, 959258.
54. Satange,R., Chuang,C.Y., Neidle,S. and Hou,M.H. (2019) Polymorphic G:G mismatches act as hotspots for inducing right-handed Z DNA by DNA intercalation. *Nucleic Acids Res.*, **47**, 8899–8912.
55. Brown,T., Hunter,W.N., Kneale,G. and Kennard,O. (1986) Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Nat. Acad. Sci. U.S.A.*, **83**, 2402–2406.
56. Shi,H., Kimsey,I.J., Gu,S., Liu,H.F., Pham,U., Schumacher,M.A. and Al-Hashimi,H.M. (2021) Revealing A-T and G-C Hoogsteen

- base pairs in stressed protein-bound duplex DNA. *Nucleic Acids Res.*, **49**, 12540–12555.
57. Zhou, H., Hintze, B.J., Kimsey, I.J., Sathyamoorthy, B., Yang, S., Richardson, J.S. and Al-Hashimi, H.M. (2015) New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res.*, **43**, 3420–3433.
  58. Hoogsteen, K. (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.*, **16**, 907–916.
  59. Aishima, J., Gitti, R.K., Noah, J.E., Gan, H.H., Schlick, T. and Wolberger, C. (2002) A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.*, **30**, 5244–5252.
  60. Nikolova, E.N., Zhou, H., Gottardo, F.L., Alvey, H.S., Kimsey, I.J. and Al-Hashimi, H.M. (2013) A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers*, **99**, 955–968.
  61. Chakraborty, D. and Wales, D.J. (2018) Energy landscape and pathways for transitions between Watson–Crick and Hoogsteen base pairing in DNA. *J. Phys. Chem. Lett.*, **9**, 229–241.
  62. Johnson, R.E., Prakash, L. and Prakash, S. (2005) Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase  $\alpha$ . *Proc. Nat. Acad. Sci. U.S.A.*, **102**, 10466–10471.
  63. 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.
  64. Golovenko, D., Bräuning, B., Vyas, P., Haran, T.E., Rozenberg, H. and Shakked, Z. (2018) New insights into the role of DNA shape on its recognition by p53 proteins. *Structure*, **26**, 1237–1250.
  65. Joerger, A.C. (2018) Extending the code of sequence readout by gene regulatory proteins: the role of Hoogsteen base pairing in p53-DNA recognition. *Structure*, **26**, 1163–1165.
  66. Nikolova, E.N., Kim, E., Wise, A.A., O'Brien, P.J., Andricioaei, I. and Al-Hashimi, H.M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–502.
  67. Frank-Kamenetskii, M.D. (2011) DNA breathes Hoogsteen. *Artif DNA PNA XNA*, **2**, 1–3.
  68. Alvey, H.S., Gottardo, F.L., Nikolova, E.N. and Al-Hashimi, H.M. (2014) Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat. Commun.*, **5**, 4786.
  69. Nikolova, E.N., Goh, G.B., Brooks, C.L. and Al-Hashimi, H.M. (2013) Characterizing the protonation State of cytosine in transient G-C Hoogsteen base pairs in duplex DNA. *J. Am. Chem. Soc.*, **135**, 6766–6769.
  70. Sathyamoorthy, B., Shi, H., Zhou, H., Xue, Y., Rangadurai, A., Merriman, D.K. and Al-Hashimi, H.M. (2017) Insights into Watson–Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m1A. *Nucleic Acids Res.*, **45**, 5586–5601.
  71. Xu, Y., McSally, J., Andricioaei, I. and Al-Hashimi, H.M. (2018) Modulation of Hoogsteen dynamics on DNA recognition. *Modulation of Hoogsteen*, **9**, 1473.
  72. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
  73. Fu, T., Liu, L., Yang, Q.-L., Wang, Y., Xu, P., Zhang, L., Liu, S., Dai, Q., Ji, Q., Xu, G.-L., et al. (2019) Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.*, **10**, 7407–7417.
  74. Wu, P.C., Tzeng, S.L., Chang, C.K., Kao, Y.F., Waring, M.J. and Hou, M.H. (2018) Cooperative recognition of T:t mismatch by echinomycin causes structural distortions in DNA duplex. *Nucleic Acids Res.*, **46**, 7396–7404.
  75. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.Refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
  76. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
  77. Jhan, C.-R., Satange, R., Wang, S.-C., Zeng, J.-Y., Horng, Y.-C., Jin, P., Neidle, S. and Hou, M.-H. (2021) Targeting the ALS/FTD-associated A-DNA kink with anthracene-based metal complex causes DNA backbone straightening and groove contraction. *Nucleic Acids Res.*, **49**, 9526–9538.
  78. Brooks, B.R., Brooks, C.L. 3rd, Mackerell, A.D. Jr, Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
  79. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. and Klein, M. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
  80. Thomsett-Scott, B. (2009) Software review of Origin 8. *J. Chem. Inf. Model.*, **49**, 2010–2010.
  81. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
  82. Chawla, M., Oliva, R., Bujnicki, J.M. and Cavallo, L. (2015) An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. *Nucleic Acids Res.*, **43**, 6714–6729.
  83. Krepl, M., Otyepka, M., Banáš, P. and Šponer, J. (2013) Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: molecular dynamics thermodynamics integration study. *J. Phys. Chem. B*, **117**, 1872–1879.
  84. Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
  85. Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. and Reid, B.R. (1983) Assignment of the non-exchangeable proton resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods. *J. Mol. Biol.*, **171**, 319–336.
  86. Ngo, T.T.M., 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.
  87. Hodges-Garcia, Y. and Hagerman, P.J. (1995) Investigation of the influence of cytosine methylation on DNA flexibility (\*). *J. Biol. Chem.*, **270**, 197–201.
  88. Furukawa, A., Walinda, E., Arita, K. and Sugase, K. (2020) Structural dynamics of double-stranded DNA with epigenome modification. *Nucleic Acids Res.*, **49**, 1152–1162.
  89. Theruvathu, J.A., Yin, Y.W., Pettitt, B.M. and Sowers, L.C. (2013) Comparison of the structural and dynamic effects of 5-methylcytosine and 5-chlorocytosine in a CpG dinucleotide sequence. *Biochemistry*, **52**, 8590–8598.
  90. Tippin, D.B. and Sundaralingam, M. (1997) Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCm5CGG) and d(CCGGGCC(Br)5CGG) in three different conformations: effects of spermine binding and methylation on the bending and condensation of A-DNA11 Edited by A. Klug. *J. Mol. Biol.*, **267**, 1171–1185.
  91. Shi, K., Pan, B., Tippin, D. and Sundaralingam, M. (2004) Structures of d(Gm5)CGm5CGCGC and d(GCGCGm5CGm5C): effects of methylation on alternating DNA octamers. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 61–65.
  92. Renciuik, D., Blacque, O., Vorlickova, M. and Spingler, B. (2013) Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.*, **41**, 9891–9900.
  93. Neidle, S. (2021) Structured waters mediate small molecule binding to G-quadruplex nucleic acids. *Pharmaceuticals (Basel, Switzerland)*, **15**, 7.

94. McQuaid, K.T., Takahashi, S., Baumgaertner, L., Cardin, D.J., Paterson, N.G., Hall, J.P., Sugimoto, N. and Cardin, C.J. (2022) Ruthenium polypyridyl complex bound to a unimolecular chair-form G-quadruplex. *J. Am. Chem. Soc.*, **144**, 5956–5964.
95. Miao, Z. and Cao, Y. (2016) Quantifying side-chain conformational variations in protein structure. *Sci. Rep.*, **6**, 37024.
96. Teng, X. and Hwang, W. (2018) Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.*, **114**, 1791–1803.
97. Mayer-Jung, C., Moras, D. and Timsit, Y. (1998) Hydration and recognition of methylated CpG steps in DNA. *EMBO J.*, **17**, 2709–2718.
98. Privé, G.G., Yanagi, K. and Dickerson, R.E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.*, **217**, 177–199.
99. Li, C., Xiong, W., Liu, X., Xiao, W., Guo, Y., Tan, J. and Li, Y. (2019) Hypomethylation at non-CpG/CpG sites in the promoter of HIF-1 $\alpha$  gene combined with enhanced H3K9Ac modification contribute to maintain higher HIF-1 $\alpha$  expression in breast cancer. *Oncogenesis*, **8**, 26.
100. Xu, Y., Manghrani, A., Liu, B., Shi, H., Pham, U., Liu, A. and Al-Hashimi, H.M. (2020) Hoogsteen base pairs increase the susceptibility of double-stranded DNA to cytotoxic damage. *J. Biol. Chem.*, **295**, 15933–15947.
101. Bohnuud, T., Beglov, D., Ngan, C.H., Zerbe, B., Hall, D.R., Brenke, R., Vajda, S., Frank-Kamenetskii, M.D. and Kozakov, D. (2012) Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Res.*, **40**, 7644–7652.