Epigenetics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/kepi20

Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation

Alexandra T P Carvalho\textsuperscript{a}, Leonor Gouveia\textsuperscript{ab}, Charan Raju Kanna\textsuperscript{a}, Sebastian K T S Wärmländer\textsuperscript{c}, Jamie A Platts\textsuperscript{d} & Shina Caroline Lynn Kamerlin\textsuperscript{a}

\textsuperscript{a} Science for Life Laboratory; Department of Cell and Molecular Biology; Uppsala University; Uppsala, Sweden
\textsuperscript{b} Department of Immunology; Genetics and Pathology; Rudbeck Laboratory; Uppsala University; Uppsala, Sweden
\textsuperscript{c} Department of Biochemistry and Biophysics; Stockholm University; Stockholm, Sweden
\textsuperscript{d} School of Chemistry; Cardiff University; Park Place, Cardiff, UK

Published online: 27 Jan 2015.

To cite this article: Alexandra T P Carvalho, Leonor Gouveia, Charan Raju Kanna, Sebastian K T S Wärmländer, Jamie A Platts & Shina Caroline Lynn Kamerlin (2014) Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation, Epigenetics, 9:12, 1604-1612, DOI: 10.4161/15592294.2014.988043

To link to this article: http://dx.doi.org/10.4161/15592294.2014.988043

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Versions of published Taylor & Francis and Routledge Open articles and Taylor & Francis and Routledge Open Select articles posted to institutional or subject repositories or any other third-party website are without warranty from Taylor & Francis of any kind, either expressed or implied, including, but not limited to, warranties of merchantability, fitness for a particular purpose, or non-infringement. Any opinions and views expressed in this article are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor & Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

It is essential that you check the license status of any given Open and Open Select article to confirm conditions of access and use.
Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation

Alexandra T P Carvalho, Leonor Gouveia, Charan Raju Kanna, Sebastian K T S Wärmländer, Jamie A Platts, and Shina Caroline Lynn Kamerlin

Abstract

We report a series of molecular dynamics (MD) simulations of up to a microsecond combined simulation time designed to probe epigenetically modified DNA sequences. More specifically, by monitoring the effects of methylation and hydroxymethylation of cytosine in different DNA sequences, we show, for the first time, that DNA epigenetic modifications change the molecule’s dynamical landscape, increasing the propensity of DNA toward different values of twist and/or roll/tilt angles (in relation to the unmodified DNA) at the modification sites. Moreover, both the extent and position of different modifications have significant effects on the amount of structural variation observed. We propose that these conformational differences, which are dependent on the sequence environment, can provide specificity for protein binding.

Introduction

Epigenetic modifications of DNA bases, such as cytosine methylation, hydroxymethylation, formylation, and carboxylation, have challenged the traditional view of the genetic system. These chemical modifications do not affect the base-pairing properties of the nucleotides, i.e., they leave the genetic code of the DNA sequence unchanged. Yet, gene expression is altered in different tissues, diseases, and developmental stages, and the modifications are passed on to offspring even through the male germ line.

The predominant (and thus far most studied) DNA modification is cytosine methylation at position C5. In mammals, such methylation occurs almost exclusively at CpG dinucleotides, and 3–6% of all cytosines in the human genome are methylated. CpG dinucleotides are mainly found on CpG islands, which are genomic regions with a high G-C content. These islands are associated with gene promoters and tend to be unmethylated, although several cases of tissue-specific and developmental stage-specific methylation of CpG islands have been reported. Methylation cytosines mainly prevail in intergenic regions, genomic repetitive elements, and highly consipuous transposable elements, such as long interspersed transposable elements (LINEs), short interspersed transposable elements (SINEs) and in endogenous retroviruses. DNA methylation is preserved after DNA replication or cell division by DNA methyltransferase 1 (DNMT1), which recognizes hemimethylated DNA and turns it into fully methylated DNA, but it can also be formed de novo by members of the DNMT3 family of enzymes.

Another interesting cytosine modification is hydroxymethylation, which was recently discovered to also be present in significant amounts in the genome. While its functions are still unclear, it appears to be involved in gene regulation, and interference with hydroxymethylated cytosines can result in a variety of co-occurrences of DNA with various modifications.
of cancers, including myeloid cancers. Even though no protein has been found to interact with hydroxymethylation sites on DNA, 10-11 translocation proteins (TET) have been identified as converters of methylcytosine to hydroxymethylcytosine.

Functionally, DNA methylation is crucial for both the developmental and adult stages of an organism. Its presence in highly repetitive sequences makes it a powerful mechanism for gene silencing, in order to maintain genome stability and integrity. In the absence of methylation in these regions, illegitimate recombination can occur, with severe mutagenic consequences for the genome, such as duplications, deletions, insertions, and translocations. Additionally, inactivation of transposable elements is of extreme importance, because these elements are also highly mutagenic and can self-replicate and be inserted in other regions of the genome, most likely disrupting essential genes for normal cellular function.

It is believed that DNA methylation leads to gene silencing by 2 different mechanisms: first, methylated CpG-binding proteins (MeCP1 and 2) recognize and bind methylated CpG dinucleotides and themselves alter transcription due to their transcription repression domain. They also recruit other factors involved in chromatin condensation. Additionally, methylated CpG nucleotides may directly inhibit the binding of transcription factors and be inserted in other regions of the genome, most likely disrupting essential genes for normal cellular function.

A common challenge when trying to model such conformational diversity is the lack of force field parameters for modified nucleotides that are compatible with generally used biomolecular force fields. For this reason, the number of molecular dynamics (MD) studies with modified DNA is limited. Nevertheless, important insight has come from studies where protein/DNA complexes were analyzed. These studies have shown that cytosine methylation destabilizes nucleosome stability by altering DNA stiffness, and another study found that methylating CpG at the major groove could enhance K50-class homeodomain PITX2 binding. However, while such studies of modified DNA in complex with proteins are clearly crucial, it is equally important to obtain more detailed insight into the effect of epigenetic modifications on the structure of the free (uncomplexed) DNA. Such studies can show if there are important differences between the normal and the modified DNA that could lead to a different recognition and hence regulatory patterns.

In the present work, we have conducted an extensive study of methylated and hydroxymethylated DNA, performing molecular dynamics (MD) simulations on DNA dodecamers and on a 20 bp long sequence both with and without modifications. The use of MD simulations to sample the conformational space of the DNA made it clear that the methylated (and hydroxymethylated) cytosines have a marked effect on the DNA structural parameters. This effect was most pronounced for symmetrically methylated CpG dinucleotides, which show different values for the twist, roll and tilt parameters near the methylation sites in relation to the unmodified DNA structure. We propose that increased flexibility and accessibility at the modification regions can have important consequences for the recognition of modified cytosines by MeCPs, transcription factors, and other elements involved in gene silencing.

**Results**

**Considerations about the methods employed**

A growing interest in the use of theoretical models to obtain a better understanding of DNA structure and its dynamics in general and, in particular, the influence of epigenetic modifications

\[ \text{Figure 1. From left to right: cytosine, 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), thymine.} \]
However, the ability of available methods to represent experimental data with structural accuracy is still unclear. There have been many elegant recent studies of DNA base pairing using fully quantum chemical (DFT) and hybrid quantum classical (QM/MM) approaches, as well as classical force field descriptions of DNA (MD simulations). In the present work, we are particularly interested in exploring how the dynamical behavior of DNA is potentially affected by epigenetic modifications. Therefore, in order to have an ensemble of structures that most correctly describes such a flexible molecule as DNA, all subsequent work was done with purely classical MD simulations. This approach allows for proper conformational sampling of the DNA strands and for concerted motions of the DNA to be studied. Moreover, when performing MD simulations, it is necessary to take into account that the unrestrained conditions that are applied during the simulation lead to a greater elasticity of the terminal part of the structure (the so-called "end-effect"). For these reasons, small oligomers, for example trinucleotides, are inadequate models to predict the reliability of the different methods or whether epigenetic modifications affect the structure of DNA.

**DNA flexibility and the differences between WT and modified DNA**

DNA is a polymorphic molecule; thus, conformational sampling is required to correctly describe its conformational flexibility. For this reason, we performed most of our analysis on a system large enough to exhibit conformational flexibility without end-effects biasing the simulation outcome. Specifically, we started with the well-studied DDD (Dickerson-Drew Dodecamer) structure, containing modifications at different cytosines, as illustrated in Figure 2. For these molecules, the initial coordinates were obtained from the reported crystal structures, except for the 3,9-DNA structure that was modeled according to the Methodology section (see also Fig. 2). We also modeled an additional sequence that was previously subjected to AFM experiments to uncover the effect of methylation on strand separation (5'-CCGAGATATCGCACCAACG-3'). The MD trajectories were then analyzed using the 3DNA software, as outlined in the Methods section. A summary of the simulations performed in this work can be found in Table S1.

Figure 3 shows the frequency of the obtained values for the base step parameters for the unmethylated DD structure and the structure methylated at the 3rd and 9th position (3,9-DNA). Note that Zgarbóva and coworkers have recently explored the
impact of base-pair fraying in molecular dynamics simulations of DNA and RNA, highlighting the challenges involved in obtaining an accurate physical description of terminal residues using contemporary force fields. Although we did not observe any compromising end-fraying effects on the timescale of our simulations, except for the standard faster base-opening, we have for safety removed the 2 base pairs (bps) at the ends (1 and 12) from the analysis. Similarly, although we show them in the graphs, we did not consider the differences in the bps 2 and 11, due to their proximity to the ends.

The central A-T regions (AA/TT, AT-AT, TT-AA) are almost identical in every parameter, between all structures. They display a smaller range of values than any of the other bp steps, meaning that they only sample a limited region of conformational space, in accordance to the stiffness previously reported for TA bps.25 However, there are differences in the steps flanking the modified base pair compared to the same bases in the WT. This is specifically seen in the twist parameter for the 3-DNA structure (Fig. 1, Table S2), which defines the overall flexibility of the oligomer, and is associated with the winding/unwinding of the helix.20 The methylated structure displays an average twist value that is ~3° lower for the 5-mCG-GC and CG-5-mCG steps comparing to the corresponding steps in the unmethylated structure. This local unwinding has been reported for other modified bases, e.g., in 2 crystal structures with N6-methylated adenine residues, these bps were unwound by roughly 2° in comparison to the unmodified DNA or in O6-alkylated guanine lesions, the guanine dimers showed strong unwinding (5 to 6°).31 For the 9-DNA structure (Fig. S2, Table S3), we observe a smaller variation in the twist value. However, there is a ~2–3° difference in the roll for the GA-T5-mC and T5-mC-GA steps. The 3,9-DNA structure (Fig. 3, Table S4) presents differences in both the twist and roll angles. It presents similar differences to the sum of the 3-DNA and the 9-DNA structures, but also shows ~2 to 3° differences in the roll of the G5-mC-5-mCG steps and differences of ~2 to 3° in the tilt for the GA-T5-mC and T5-mC-GA.

Regarding the local base pair parameters, most of them show no differences between the WT and modified DNA on this timescale, except for the buckle and the propeller twist, which show small differences near the modification sites. The differences are smaller than those observed for the base step parameters. Figs. S3-S5 and Tables S5-S7 show the obtained values of the local base parameters for 3’100ns simulations of the WT and methylated sequences at position 3, 9 and both 3 and 9 (see Fig. 2). Also, Figs. S6-S9 and Tables S8-S11 show the values for the base step parameters and local parameters for the hydroxymethylated structures at position 3 and 9.

The 3-DNA and 9-DNA structures are different with regard to the twist angle. Specifically, the 3-DNA structure shows significant local unwinding near the methylation position, whereas the 9-DNA structure behaves mostly as the unmodified structure with regard to this parameter (Fig. S1 and S2, and Tables S2 and S3). The atomic level explanation for this relies on the relative position of the methyl groups at the major groove of DNA. The 3-DNA has the methyl group of the modified nucleotide in one strand, whereas the methyl groups of the thymine bases 7 and 8 are on the other strand. Although these residues are too far away to directly interact, there is a change in the hydration patterns that leads to localized untwisting of the DNA, since the methyl group in the modified base pair creates a hydrophobic region on the DNA (the corresponding RDF of which is shown in Fig. 4), followed by a more hydrated region (Fig. 5). Note that the methyl hydrogens are still able to interact with water molecules; however the water molecules are further away from the base. This is visible in the RDF of the H41 hydrogen (one of
the amine hydrogen atoms), which in the unmodified structure has a much higher probability to interact with a water molecule than for the methylated sequence. The DNA molecule seems to unwind in order to accommodate the change in solvation.

This effect is not observed for the 9-DNA structure, since both the methylated base and the 2 thymine bases are in the same strand, and then the twisting of the structure would not lead to any proximity between the hydrophobic regions (Fig. 2). The opposite scenario occurs with regard to the roll and tilt parameters, where we observe differences in the 9-DNA structure and in 3,9-DNA structures, but not in the 3-DNA structure. This local bending can also be a consequence of the proximity of the methyl groups of the modified cytosine bases and of the thymine bases lying in the major groove (Fig. 2).

The central part of the dodecamer is an A-tract with 4 A/T base pairs, a sequence...
that has its own intrinsic properties including a fully methylated major groove.\textsuperscript{32,33} We have previously shown that continuous major groove methylation affects both the structure and dynamics of dsDNA by improving base stacking and probably also by modulating water binding.\textsuperscript{34-36} Thus, it might be of significance that methylation of d(CGCGAATTCGCG)\textsubscript{2} at position 9 creates a 6 bases long spine of major groove methyl groups, that double-methylation at positions 3 and 9 provides 8 consecutive methyl groups, and that the methyl group afforded by single methylation at position 3 remains isolated (Fig. 2).

For the hydroxymethylated structures, we observe similar profiles as for the methylated structures. Furthermore, our simulations also allow us to probe the preferred orientation of the hydroxymethyl group: DFT predicts a slight preference for the OH group to point in 3' over 5' and an optimized C6-C5-C5A-O5 torsion angle of 118.4\textdegree/C14, while our MD simulations show this angle to vary between 85 and 120\textdegree over 100 ns of simulation. This is in good agreement with the experimental and theoretical results of Renciuk et al.,\textsuperscript{12} who reported values between 72 and 133\textdegree using X-ray diffraction methods.

These results indicate that DNA modifications can increase DNA flexibility and alter the twisting angle of the structure if the methylation is in a different strand than a neighboring thymine. Large twist fluctuations for CG-CG steps have been reported before\textsuperscript{23} and have been shown to be sequence dependent. Finally, if the methylation is in the same strand as a neighboring thymine we observe differences in roll and tilt, which are related to DNA bending. From our analysis it seems that these small chemical modifications (methylation and hydroxymethylation of cytosines) lead to increased structural diversity in the roll, tilt and twist parameters, which are related to unwinding and bending of the DNA. This in principle can modulate the affinity of the DNA to proteins and hence have a role in protein-DNA recognition.

For the 20 bp structure we found that methylation leads again to modifications in the twist and roll/tilt parameters near the modified sites (Fig. 6 and Table S12). However, now the pattern is more complicated than before, since now there are more neighboring thymine residues, which leads to changes between the modified and unmodified sequences further away from the central methylated CpG. Direct comparison of these sequences with protein DNA complexes is not possible, since in all the crystals of protein-methylated DNA complexes the methylated nucleotide is flipped.\textsuperscript{37-39} However, indirect readout or shape (topography) recognition is a well-known mechanism for DNA recognition.\textsuperscript{40} This was further corroborated by Hansen et al., who showed that DNA shape is under evolutionary selection pressure.\textsuperscript{40,41} Consequently, it is feasible that the differences in DNA shape upon methylation observed by us can influence protein-DNA recognition and help to clarify the structural and mechanistic details of epigenetics.

**Discussion**

The regulation of gene expression by DNA methylation is of great importance, since it is required for maintaining genomic stability in cells. Changes in methylation patterns can give rise to dreadful consequences for the cell (and, therefore by extension, the organism), and are involved in many human diseases, including cancer. Studies of DNA methylation have provided insightful information about the factors underlying the establishment and maintenance of methylation in mammalian cells,\textsuperscript{42} but corresponding work aimed at obtaining a fundamental understanding of how the intrinsic behavior of DNA is influenced by such modifications has been limited.
The use of theoretical methods to describe complex biomolecules such as DNA is becoming increasingly common, with theory providing further insight into the interpretation of the experimental data. One example of this is the work of Schulten and co-workers on DNA strand separation. These authors combined single molecule force microscopy experiments with steered MD simulations and showed that cytosine methylation and hydroxymethylation alter the mechanical properties of DNA, thus facilitating strand separation. However, although they identified this effect, they did not determine its underlying mechanism. Another important example is a recent study of methylation in CpG-rich sequences, which showed that methylation lowers the energy difference between the B- and Z-DNA conformations for these highly repetitive sequences. However, the authors again did not find any apparent significant differences in structure that accounted for this energy difference.

In the present work, we report the structural dynamics of 5-methylcytosine and 5-hydroxymethylcytosine DNA sequences. Overall, the employment of force field and molecular dynamics approaches appears to provide a better representation of the experimental data to when compared to a static description provided by hybrid QM/MM approaches when compared to a static description. The flexibility and deformability of native DNA has been extensively studied, and it is well established that both features are sequence-dependent, with base composition and the presence of specific sequences substantially contributing to the malleability of the double helix. In the present work, the different conformations that DNA can adopt were evaluated by measuring relevant base pair parameters during our simulations. Previous work has shown that the AA/TT and GA/TC steps are considered to be the ones with less variability. The obtained data from our MD simulations concerning all the modified structures are consistent with this observation as far as the AA/TT step is concerned, which shows no variability at all. However, the authors again did not find any apparent significant differences in structure that accounted for this energy difference.

The two consecutive methylations affected the step conformations in the same way as the combination of the single methylations, except for the roll parameter. In this last case, the cytosine modifications do in fact have an effect on the conformational state of the steps as well as on the overall structure.

These differences in the conformational flexibility of DNA were even more pronounced when we created a symmetrically methylated CpG dinucleotide-containing structure. From our point of view, this structure was of extreme importance, since it contains a suitable amount of methylated cytosines to properly represent cellular patterns of methylation. The two consecutive methylations affected the step conformations in the same way as the combination of the single methylations, except for the roll parameter. In this last case, the differences due to the unmodified structure are more pronounced than in the 9-DNA structure. We do not believe that the observed differences are due to end-effects. Otherwise, we would not observe consistent results for the different types of structural modifications examined (i.e., methylated and hydroxymethylated DNA).

The comparison of results between the 2 doubly modified structures, with different positions of modifications, confirms that DNA conformational flexibility depends on the presence of particular sequences, and that the neighboring steps influence the conformation adopted by each step. Thus, static structures do not provide a good description of such a flexible molecule as DNA, and a proper description of the behavior of such systems must take their flexibility into account. Techniques that report static structures, such as purely QM optimization or X-ray crystallography may therefore miss key differences in the base-pair helical parameters between non-modified and modified structures. In this study, we show unprecedented results demonstrating that modified structures have the propensity to adopt different conformations, which can hypothetically have an impact on recognition of modified cytosines by important proteins, such as DNMTs and MeCPs. Finally, this augmentation in flexibility may be an aid for chromatin condensation, which is a process also connected to methylated cytosines and gene silencing.

Methods

Initial system setup

Atomic coordinates of the wild-type, methylated and hydroxymethylated DNA dodecamers examined in this work, based on Dickerson’s well-studied self-complementary d(5'-CCGGAATTCGG-3')2 sequence, were obtained from X-ray structures deposited in the Protein Data Bank (PDB IDs: 1BNA, 4GJU, 4GLG, 4GLH and 4GLC). The 4GJU structure is methylated in the position 3 of each strand and the 4GLG structure is methylated in each position 9. We also constructed an additional structure methylated in positions 3 and 9 (4 methylations), this structure mimics 2 symmetrical methylated CpG dinucleotides, as it occurs in the cells. For simplicity, in the text, we refer to the 4GJU structure as 3-DNA, and 4GLG structure as 9-DNA, and the combined structure 3,9-DNA. We also analyzed an additional sequence (5'-CCGAGATATCCGGACACAACG-3') and the corresponding methylated one (5'-CCGAGATATCmCGACACAACG-3'), which were initially modeled using the X3DNA program.

Calculation of point charges

All QM calculations were performed with the Gaussian09 simulation package, with HF as well as the 6–31G(d,p) basis set. Charges for the methylated and hydroxymethylated bases were obtained using the restrained electrostatic potentials method (RESP), in accordance with the traditional AMBER charge derivation methodology (Tables S13 and S14).

Molecular dynamics simulations

All molecular dynamics (MD) simulations were performed using the GROMACS simulation package and the amber parmbsc0 force field. This force field was specifically chosen because it has an improved description of the α/γ-concerted rotation in nucleic acids compared to simple parm99. For a more
detailed discussion on available force fields for nucleic acids please consult Sponer et al.61 The DNA models were enclosed in a cubic box with a volume of 68.47 nm³ (2.403 nm × 2.403 nm × 2.403 nm) containing TIP3P water molecules62, as well as Na⁺ and Cl⁻ counter ions to create a neutral electrostatic environment. Electrostatic interactions were accounted for by using the fast Smooth Particle-Mesh Ewald (SPME) method,63 with a 10Å cutoff being employed to both electrostatic and van der Waals interactions.

We prepared our system by performing an initial brief energy minimization with 200 steps of the full simulation system, followed by 2 500ps equilibration runs. The first equilibration run was performed in a canonical ensemble with the standard velocity rescaling thermostat,64 and the second equilibration run was conducted in the isothermal-isobaric ensemble using the Berendsen thermostat.65 The temperature and pressure coupling during the simulations performed on the Triolith and Tintin clusters at the Swedish National Infrastructure for Computing (SNIC) for generous allocation of computer time (SNIC grant 025/12–10). This work was supported by calculations performed on the Triolith and Tintin clusters at the National Supercomputing Center in Linköping and UPPMAX in Uppsala respectively. JAP is grateful for access to Advanced Research Computing at Cardiff (ARCA) facilities.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We would like to thank the Swedish National Infrastructure for Computing (SNIC) for generous allocation of computer time (SNIC grant 025/12–10). This work was supported by calculations performed on the Triolith and Tintin clusters at the National Supercomputing Center in Linköping and UPPMAX in Uppsala respectively. JAP is grateful for access to Advanced Research Computing at Cardiff (ARCA) facilities.

Funding

This work was supported by the Swedish Research Council (Vetenskapsrådet, 2010–2006) and funding from the Sven and Ebba Christina Hagbergs Foundation.


