Current Topics in Medicinal Chemistry, 2020, *20*, 901-909

REVIEW ARTICLE

Towards a Better Understanding of Computational Models for Predicting DNA Methylation Effects at the Molecular Level

Nathanael K. Proctor[1], Tugba Ertan-Bolelli[2,3], Kayhan Bolelli[2,3,4], Ethan W. Taylor[1], Norman H.L. Chiu[1] and J. Phillip Bowen[2,*]

[1]*Department of Chemistry & Biochemistry, University of North Carolina Greensboro, Greensboro, NC 27402, USA;* [2]*Department of Pharmaceutical Sciences, College of Pharmacy, Mercer University, Atlanta, GA 30341, USA;* [3]*Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ankara University, Ankara, TR 06560, Turkey;* [4]*LumiLabs LLC, Ulus, Ankara, TR 06050, Turkey*

ARTICLE HISTORY

Received: December 15, 2019
Revised: December 24, 2019
Accepted: December 24, 2019

DOI:
10.2174/1568026620666200226110019

Abstract: Human DNA is a very sensitive macromolecule and slight changes in the structure of DNA can have disastrous effects on the organism. When nucleotides are modified, or changed, the resulting DNA sequence can lose its information, if it is part of a gene, or it can become a problem for replication and repair. Human cells can regulate themselves by using a process known as DNA methylation. This methylation is vitally important in cell differentiation and expression of genes. When the methylation is uncontrolled, however, or does not occur in the right place, serious pathophysiological consequences may result. Excess methylation causes changes in the conformation of the DNA double helix. The secondary structure of DNA is highly dependent upon the sequence. Therefore, if the sequence changes slightly the secondary structure can change as well. These slight changes will then cause the double-stranded DNA to be more open and available in some places where large adductions can come in and react with the DNA base pairs. Computer models have been used to simulate a variety of biological processes including protein function and binding, and there is a growing body of evidence that *in silico* methods can shed light on DNA methylation. Understanding the anomeric effect that contributes to the structural and conformational flexibility of furanose rings through a combination of quantum mechanical and experimental studies is critical for successful molecular dynamic simulations.

Keywords: DNA, DNA methylation, nucleosides, nucleotides, molecular mechanics (MM), molecular dynamics (MD), MD simulations, computational chemistry, quantum mechanics (QM), density functional theory (DFT), QM/MM, *ab initio* calculations, anomeric effect, conformations, pseudorotation.

1. INTRODUCTION

Human cells can regulate themselves by using a process known as DNA methylation. This methylation is vitally important in cell differentiation and expression of genes [1]. In the context of biological studies, DNA methylation is one of many modifications that can occur in genomic DNA of both prokaryotes and eukaryotes [2, 3]. Specifically, DNA methylation refers to specific enzymatic reactions in which a methyl group is transferred from S-adenosyl methionine to a specific location of 2'-deoxyribonucleotide [4]. In the case of cytosine, 5-methylcytosine (m5C) is the predominant modification. In prokaryotes, the presence of m5C can provide a defense mechanism that resists the activity of restrictive enzymes. Many m5C blocks are found at the CpG sites within a specific genome, where a cytosine is immediately in front of a guanine [5]. The methylation of cytosine is further complicated by the fact that m5C may undergo additional modifications. In other words, m5C can serve as a precursor for other types of DNA modification, which include 5-hydroxymethylcytosine (5-hmC) [6], 5-formylcytosine (5-fC), or 5-carboxylcytosine (5-caC) [2, 6, 7]. Although these modifications at the 5 position of cytidine do not interfere with the Watson-Crick base pairing between the modified cytidine and guanine [8], each of these modifications including m5C is expected to induce conformational changes in the double DNA helix [9], which in turn can affect the interactions between DNA and histones [10]. As a result, DNA methylation can potentially alter the outcome of gene expression, as well as the replication of genomic DNA.

When methylation is uncontrolled or does not occur in the right place, there can be serious consequences. Clinically, DNA methylation has been linked to several pathologic conditions (*e.g.,* cancer [11], diabetes mellitus (DM) [12], asthma [13, 14]. Although environmental factors and genetic variations are vital to the development and progression of DM type 2, according to recent studies DNA methylation plays a key role in the pathogenesis of DM type 2 by affecting the body's resistance to insulin and its secretion of pan-

*Address correspondence to this author at the Department of Pharmaceutical Sciences, College of Pharmacy, Mercer University, Atlanta, GA 30341, USA; E-mail: bowen_jp@mercer.edu

creatic β cells [12]. Epigenetic and genetic processes such as DNA methylation and miRNA expression have been implicated in asthma pathogenesis. Actually, in asthmatic blood cells, DNA methylation is altered and it can be a biomarker of atopy [13, 14]. Overall, the knowledge and understanding on the methylation of the human genome have advanced our study on etiology and assisted the development of new pharmacologic therapies [15]. An example is N7-methyl-2′-deoxyguanosine (MdG) which is a DNA alkylating agent. In a recent study, MdG in nucleosome core particles (NCPs) were examined. A basic site formation from MdG was suppressed in NCPs. Besides, MdG and histone proteins form cross-links that cause harmful DNA damages and may eventually lead to cytotoxicity [16]. Nevertheless, there are still gaps in our knowledge of DNA methylation and their relationships with environmental factors and other genetic elements [3].

2. COMPUTATIONAL STUDIES

Recently, computer models have been used to simulate a wide variety of biological processes, including protein function and ligand-target binding. Human DNA is a very sensitive macromolecule. Slight alterations to the structure of DNA can have short- and long-term disastrous consequences. When nucleotides are modified or changed the resulting DNA sequence can lose its information, if it is part of a gene, or this can become a problem for replication and repair [17]. In one case after an *N*-acetyl-2-aminofluorene or an *N*-2-aminofluorene was added to a guanine, the DNA sequence was changed in very specific ways [18]. It has long been known that DNA methylation is an important part of the cell cycle, and healthy cells have many methylated base pairs. Sixty percent of cytosine residues in CpG islands contain a methyl (*via* methylation) at the 5th position of the cytosine [19]. This can lead to problems since these CpG islands contain about 75% of the promoters for genes [20]. The problems occur where there is excess-methylation and the DNA becomes subject to different types of adduction (*e.g.,* the carcinogenic diol epoxide metabolite anti-benzo[*a*]pyrene diol epoxide, BPDE). BPDE is an adduct that can occur at multiple DNA bases. For example, it can occur on both the 3′ and 5′ sides of the adenine base although one side is more stable than the other [21]. Once the base has this adduction on it, changes occur in the stability of the DNA double helix, and it is possible that this change in stability causes the base to be recognized by a repair enzyme.

For the analysis of BPDE-DNA adducts, several analytical methods have been used, including immunoassay, ^{32}P-postlabeling, and LC-MS/MS-based methods [22-24]. Among these analytical methods, the best one is the LC-MS/MS. For the first time, the quantification of BPDE-dG adducts in human umbilical cord blood samples with the LC-MS/MS method was used [25], and the method that was developed has a sensitivity which can be applied to use human cord blood cell DNA in a quantity as low as ~10 mg of DNA. In an earlier mass spectrometry study, Paul Vouros and his group studied the effect of DNA methylation on a double-stranded DNA helix (5′-ACCCG5CG7TCC G11CG13C-3′/5′-GCGCGGGCGCGGGT-3′) [26]. In the study, they methylated several cytosine residues on one strand, known as strand 1, and incubated the DNA with the known adduct BPDE. Then, the same setup was repeated with methylation of cytosine residues on the second strand, strand 2, and finally with both strands methylated on the specific cytosine residues. The next step was to determine if the BPDE adduct occurred and if it occurred more often in one place relative to the others. The BPDE adduct increases the mass of the DNA, and therefore a method of mass spectrometry using an online nano LC-MS/MS was used to determine the adduct formation. It is known that a BPDE adduct binds to DNA at the N2 position of the guanine residue [27]. When the results were analyzed from the mass spectrometry data, it was found that adduction occurred on all of the guanine residues with the guanine at the fifth position, G5, which is the most commonly observed BPDE adduction site. With only one strand methylated, the other guanine positions showed that G11 had greater adduct formation than G13, which in turn was greater than G7. When both of the strands were methylated, the sequence of adduction changes such that G5 had greater adduction than G7, which was greater than G11, which, in turn, was greater than G13. Therefore, with excess methylation a difference was seen in the trend of BPDE adduction. It was concluded that cytosine methylation significantly increases the chances of BPDE binding.

It is believed that an excess of methylation causes changes in the conformation of the DNA double helix. A study that was done previously showed that the adduct 4-OHEN forms at certain cytosine residues due to both the sequence and the secondary structure of DNA [28]. The secondary structure of the DNA is highly dependent upon the sequence. Therefore, if the sequence changes slightly, the secondary structure can change as well. These slight changes will then cause the double-stranded DNA to be more open and available in some places where large adductions can react with the DNA base pairs. Once there is a large adduct in the DNA helix, other conformational changes can occur that have the possibility to cause major problems in the cell. For example, the gene expression in a particular region may become repressed, the DNA might not be able to be replicate properly, and the gene might not be expressed. In one study, it was found that the major cause of the P53 tumor formation in lung cancer is not caused by an endogenous pathway, but instead, methylation of the CpG sites of the gene is believed to cause the chemical carcinogen adduct, BPDE, to occur more often [29]. Therefore, in many cases, the over-methylation of DNA can lead to diseases.

The goal of computational chemistry is to predict the behavior of molecular structures using quantum mechanics (QM), molecular mechanics (MM), and/or molecular dynamics (MD) simulations prior to experimentation. When applying these *in silico* methods to a biological environment, many factors influence the results of the simulated dynamics. Some of the advances and remaining problems that still exist are discussed by Cheatham and Young [30]. In their discussion, they point out that great advances have been made in the area of ion interaction. Ion interaction is important for proper DNA MD because the negative charges on the backbone of the DNA will repel each other and lead to splitting of the DNA strands. Usually, Na^+ ions are used to interact with the backbone, and a common set up is to place the sodium ions about 6 Å away from the backbone in the groove

[31]. The positive charges on the sodium ions will interact and balance with the negative ions of the phosphate groups on the DNA backbone. In another study, the impact of different types of ions (Mg^{2+} or Na^{+}) and variations of RNA and the ionic strength on structural stability were tested experimentally. According to this study, Mg^{2+} was found to conserve the experimental structure better than Na^{+} and, they can be reproduced with reasonable accuracy where experimental ion positions are available [32].

Another successful outcome of a computational approach is the ability of models to allow DNA to bend. Large DNA molecules will have a curvature, and presently some MD simulations can reproduce this effect. One final success that was mentioned was the ability to use MD on varied DNA structures. Computer-based modeling of DNA is now able to demonstrate each of the different conformations of DNA. In living cells, DNA can adopt an A-form, B-form, or Z-form. The differences between these conformations are due to the puckering of the sugar ring. Both the A- and B- forms of DNA have a right-hand helix conformation; whereas, the Z-form DNA is in a left-handed helix. Molecular modeling can be used to build these different forms of DNA. Even with all the advances that are seen, there are still, nevertheless, some problems that arise when running MD simulations. The first of these is a conformational sampling. When a dynamics simulation runs, the desired result is the one with the lowest energy trajectory or tracing the path of movement in a low energy conformer. It has been shown that the molecule might reach a local minimum and get stuck in that conformation even though it is not representative of what is really happening [33]. An analogy can be made to a valley between two mountains. At the bottom of the valley is the lowest energy conformer. When the conformers are checked for energy any conformer with energy higher than the previous conformer is discarded. Therefore, the last result will hopefully be the one that is the lowest in energy. However, perhaps there is a small perturbation in the energy surface that behaves like the energy well. If a conformer is found to be in that energy well, then it will be at a local minimum. The local minima *vs.* global minimum remains a challenging problem in MD simulations and conformational searching.

The results of energy calculations with MM and MD nucleic acid modeling remain problematic. Unlike QM, the energy equations are a function of the different force fields that are used. Each force field has different equations and parameter sets with strengths and weaknesses. Some force fields will allow the study of different types of DNA, such as the Hoogsteen base pairing that is not commonly seen in the majority of DNA but that still occurs in native DNA [34]. Another issue that has fueled concerns about DNA calculations with MM and MD is the nucleoside anomeric effect. Nucleoside sugars have different conformations, and there is an equilibrium between the North and South conformations. The anomeric effect is due to the lone pair electrons *anti* to electronegative heteroatoms. The DNA furanose contains oxygen in the ring with an aromatic nitrogen of the base, either a purine or a pyrimidine, bonded to the anomeric carbon. These two atoms lead to the anomeric effect occurring in the sugar which will affect the puckering of the sugar ring. There are two conformations, the East and West, which are the barriers that must be passed through to convert from North to South or *vice versa*. One molecular modeling study of DNA showed that the eastern barrier is too low, which allows for interconversion between the two conformations much easier than it should be [35]. Therefore, the representation of MD with DNA would not be true to reality.

The development of solvent models for biomolecular simulations continues to be a challenge. In the human cell, all of the biomolecules are in an aqueous environment rich with ions. Therefore, in order to have proper MD simulations that mimic nature, the biomolecules must be solvated. Problems arise when the DNA double helix becomes solvated. The computational cost (*i.e.,* CPU time) increases significantly when large portions of nucleic acids are solvated. This refers to the amount of resources used by the computer. Cheatham and Young [30] discuss two ways that help avoid these problems and that is by using an implicit solvation model. One potential approach is the Born methodology [36], and the second is to use a Poisson-Boltzmann method [37]. These two implicit solvation methods have helped cut down on the cost of running an MD situation. Another problem arises with the lack of structural water. In native DNA, water will align so that the hydrogen is in the backbone of the DNA. In simulations, this is still a step that is hard to accomplish, although some force fields and parameters do allow this to occur [34]. If these problems are not addressed the resulting simulations might then show bends or twists that are not representative of the true nature of DNA. To achieve neutrality in the simulation systems, current MD simulations use explicit solvent models with the addition of counter ions (K^{+} or Na^{+}). This method is a high computational expense but more accurate. The umbrella sampling MD has revealed interesting mechanisms involved in bending and base pair openings that may lead to local distortions involved in DNA repairing and epigenetic imprinting. The folding/ unfolding of small DNA molecules can be studied within the μs time scale [38, 39]. One ubiquitous problem in MD simulations is the question of charge-charge cutoffs. This issue has been solved by using the Ewald Particle-Mesh method [40].

There are many different sets of MM parameters that can be used to run a simulation. The objective is to apply a set of parameters that will give the best results for the specific problem. MMFF94s simulation parameters are different than the original MMFF94 parameters. One difference between the two models is that the MMFF94s parameters change the out-of-plane bending with planar geometries using nitrogen [41]. A study was carried out by Halgren that compared the MMFF94, MMFF94s, CFF95, CVFF, MSI CHARMm, AMBER*, OPLS*, MM2*, and MM3* force fields in determining conformational energies of a large set of known molecules [42]. The findings showed that the MMFF94 and MMFF94s force fields had the best results and were the most consistent; however, there were problems with these as well. This should be expected since no computer-based modeling scheme or mathematical model is perfect. The MM2, MM3, and MM4 series of programs have largely been replaced with more robust implementations that can handle more diverse functional groups of biological interest. Moreover, the application of MM2, MM3, and MM4 to macromolecular calculations was never widespread. Recently, a new united-atoms GROMOS force parameter set was designed and vali-

dated for furanose-based carbohydrates within an aqueous environment [43]. Validation involved comparing the MM results with QM calculations and available experimental data. The newly developed set of parameters was able to reproduce conformational equilibria based on the anomeric effect within the furanose ring, as well as the relative free energies of anomers, hydroxymethyl rotamers, and glycosidic linkage conformers.

QM has served as the basis for many studies of the anomeric effect from structural to reactivity effects. The calculations of nucleosides and nucleotides are much more cost-effective. Using the Hartree-Fock level of theory with the standard 6-31G(d) basis set the classical hydrogen bonding patterns and electrostatic potential map are easily generated with the Spartan '18 molecular modeling software [44] (Fig. **1**). Because of the synthetic and biological importance of pyranose rings, a number of computational studies have been carried out on the anomeric effect, including reactivity [45]. One recent summary of the anomeric effect in pyranose ring systems has wide implications [46].

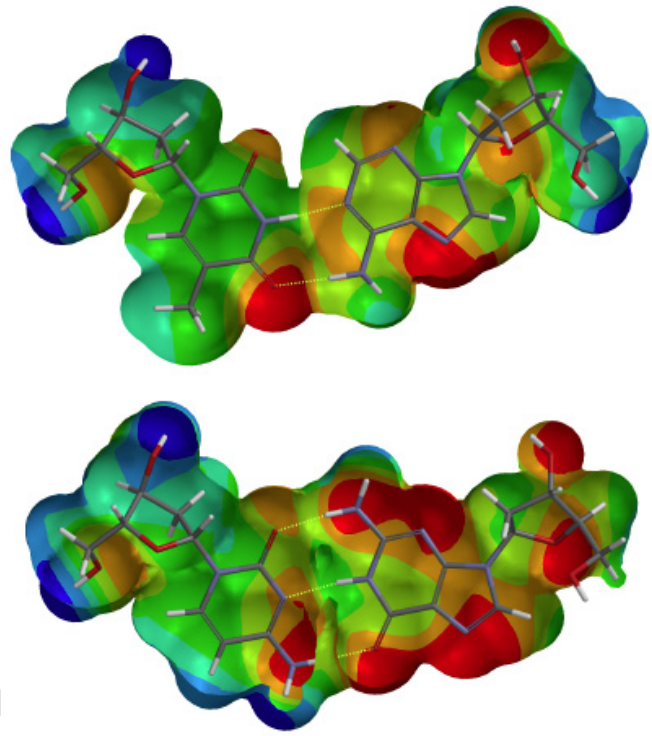

Fig. (1). The classical hydrogen bonding arrangement between base pairs in nucleosides are displayed along with an electrostatic potential map. The structures were generated using the Spartan '18 molecular modeling software [44] at the Hartree-Fock level of theory with the standard 6-31G(d) basis set. The electrostatic potential is represented by a spectrum of colors. The red color represents regions of high electron density while the blue color represents regions of low electron density.

The combination of QM and MM is an attractive approach for studying macromolecular systems. Recent work by Aranda *et al.* focused on computational studies of M.HhaI, a DNA methyltransferase from *Haemophilus haemolyticus*, which catalyzes the transfer of methyl to the C5 position of a cytosine from S-adenosyl-L-methionine (SAM) [47]. Although hydroxide ions have been proposed as the source for abstracting the proton from the C5 position, the calculations in this study indicate that the free energy required for absraction is too high. The calculations reveal that a water molecule can remove an additional proton after the cytosine has been activated *via* a proton transfer from Glu119 to the N3 position. The theoretical studies appear to resolve two issues (*i.e.,* the base and role of Glu119) that have been debated.

Also, when modeling double-stranded DNA, it is important to remember how DNA behaves in living cells. There is a breathing of DNA that occurs naturally when the two strands come apart and then hybridize back together. Studies have been carried out that show how well computers can model this behavior coupled with *in vitro* experiments [48]. In these studies, it was found that the time-scale for reliable results when using MD with DNA is less than nanoseconds. The reason for the unreliability is that the double-stranded DNA helix will fall apart after a period of time close to the nanosecond range. A newer study has been carried out and showed that a 5 nanosecond MD simulation has been observed that did not fall apart [49].

As can be seen, significant advances have been made in the area of computer modeling. There are still problems that are being addressed, and adjustments are being made constantly in order to make MD simulations reflect experimental observations. MD simulations for double helix DNA have been a challenging area, but there have been vast improvements in certain areas that allow us to have confidence in obtaining reliable results when trying to replicate experimental work that has already been carried out.

3. FUNDAMENTAL CONSIDERATIONS FOR MOLECULAR MODELING OF NUCLEIC ACIDS, NUCLEOTIDES, AND NUCLEOSIDES

At the most basic level of nucleic acid structure (the monomers of the biopolymer), a nucleo*side* is a purine or pyrimidine base joined by an N-glycosidic bond to the C1' carbon of a β-D-ribose sugar in its furanose (cyclic) form, whereas a nucleo*tide* is a 5'-mono-, di- or triphosphate of a nucleoside. An oligonucleotide (of the type found in natural DNA and RNA) is a 5',3'-linked phosphodiester nucleotide polymer. As pointed out early on Sundaralingam [50, 51], a remarkably small number of structural descriptors are sufficient to fully determine the conformation of nucleic acids at the monomer level. For an individual nucleoside (Fig. **2**), the conformation can be described by just three major parameters, if furanose ring puckering is assumed to be of constant amplitude.

The base orientation is described by the glycosidic angle χ (the torsional angle for the bond from the furanose C1' to the base N1), which is *syn* when the bulkier portion of the base (*i.e.* pyrimidine C2 carbonyl or purine N3) is over the sugar ring, *anti* when the carbonyl or purine 6-membered ring is opposite the sugar ring (Fig. **2**).

The C4'-C5' torsional angle, γ (Fig. **2**), which has three rotamers (Fig. **3**), either *+synclinal (+sc)*, *-sycnclinal (-sc)* or *antiperiplanar (ap)*, forms part of the nucleic acid backbone

and determines the orientation of O5' with respect to both C3' and the furanose ring oxygen O4', which influences the rotamer population of O5' *via* the *gauche* effect [52].

The *pseudorotational phase angle*, P (Fig. **4**) is not an actual angle at all, but rather a degree on a cycle that serves as a convention to describe the pathway by which permitted conformations of the 5-membered furanose ring can interconvert [53].

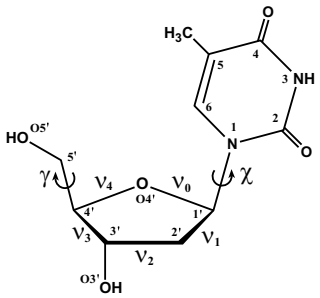

Fig. (2). Atom numbering scheme and rotatable bond naming conventions for nucleosides, shown for thymidine. Here, the glycosidic angle χ is shown in the *anti* range (pyrimidine carbonyl projecting away from the furanose ring oxygen). The C4'-C5' torsional angle γ, which determines the position of O5', is shown in the *ap* orientation (*anti* to C3'), placing O5' *gauche* relative to the ring oxygen O4' (Fig. **3**). The endocyclic torsion angles ν_0 through ν_4 determine the specific type and amplitude of furanose ring puckering.

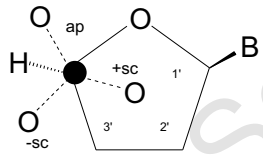

Fig. (3). The three possible rotamers of the 5'-CH_2OH group produced by varying the torsion angle γ are shown, *via* a projection down the C5'-C4' bond. The preferred *ap* and *+sc* conformers (defined relative to C3', along the DNA backbone) are both *gauche* to the ring oxygen O4', and are more commonly observed than *-sc*.

P is dependent upon the set of 5 endocyclic torsional angles, ν_0 through ν_4 (Fig. **2**), which are related in a complex fashion, due to constraints imposed by the 5-membered ring. The 20 possible different envelope and twist conformations of the furanose ring interconvert *via* the pseudorotation pathway while avoiding the unfavorable planar conformation (which is forbidden because it would necessitate fully eclipsing interactions of all the substituents on the furanose ring carbons). The beauty of Sundaralingam's concept is that, assuming that there is a constant amplitude of puckering, it permits the furanose ring conformation to be described by a single parameter, P, ranging from 0-360 [53, 54]. The convention for naming the state of the ring pucker in nucleosides is that, relative to the plane of the furanose defined by 3 (for twist conformations) or 4 (for envelopes) in-plane ring atoms, when an atom is puckering towards the same side as the base and C4', it is described as *endo*; if puckering towards the "outside" opposite the base and C4', it is described as *exo*. Then, with *endo* or *exo* determined by super- or subscript, respectively, these conformations can be abbreviated as, for example, ^{2}E, a 2'-*endo* envelope, or $^2T_3$, a 2' *endo*-3'-*exo* twist, both of which are S-type conformations (Fig. **3**). The sub- and superscripted numbers, or O for the ring oxygen, are sometimes both placed stacked on the left side of the E or T (Fig. **4**).

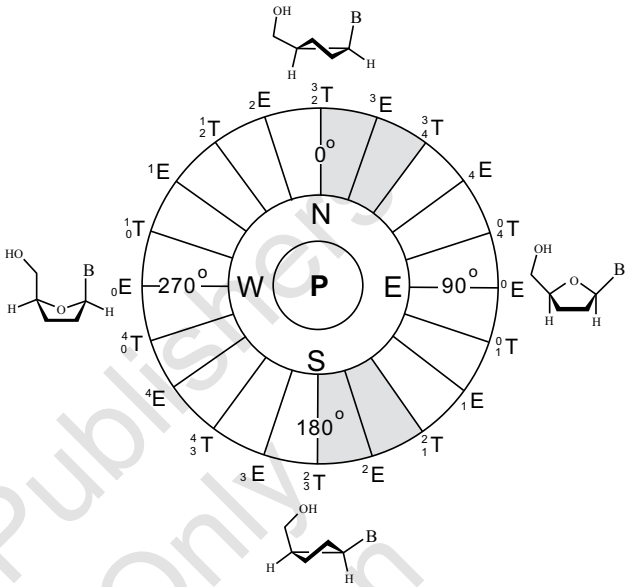

Fig. (4). The pseudorotation cycle: a pathway for interconversion between the various envelope (E) and twist (T) conformations of the furanose ring, avoiding the strained planar conformation. P, the pseudorotational phase angle. The typical N (P = 0°) and S type (P = 180°) furanose twist conformations, corresponding to energy minima, are shown above and below the figure. The far E and W conformations shown, which are both envelopes with O5' out of the plane, present barriers to interconversion between N and S types, but the W type presents a much higher barrier due to the steric clash from the diaxial positioning of the bulky ring substituents (CH_2OH and the base B). Relative to the plane formed by the other three (for twist conformations) or four (for envelope conformations) furanose ring atoms, an out-of-plane atom is *endo* (superscript) if it is on the same side of the plane as C5' and the base, or *exo* (subscript) if it puckering towards the opposite side. Thus, ^{2}E = C2'-*endo*, an S-type envelope conformation at P = 162°.

Thus, the values of the two exocyclic bonds χ and γ, along with the furanose pucker as described by P, are sufficient to fully describe the 3D state of a nucleoside monomer, subject to the added constraint of a specific MM force field. Different force fields will give slightly varying degrees of puckering amplitude, depending on the specific torsional, van der Wäals and other parameters that are invoked. But for practical purposes, assuming adequate parameterization and appropriate potential energy terms, most of the force fields in common use for biopolymer modeling should give essentially the same structure if minimized at identical values of χ, γ and P.

For oligonucleotides, as shown for a unit of RNA (Fig. **5**), the backbone conformational state is fully defined by the set of torsion angles α, β, γ, δ, ε and ζ, starting from the bond between P5' and O5' (α) and going through the sugar backbone carbon atoms towards O3' and its bond (ζ) to the P

atom of the next nucleotide unit. Note that the γ backbone angle is the same as the exocyclic γ angle defined above for isolated nucleosides. In canonical nucleic acid structures, relatively few conformational types or torsional ranges are observed for these descriptors, out of a large number of theoretical possibilities. We will briefly review these preferences, as follows:

Orientations of the glycosidic angle χ tend toward the extremes of either s*yn* or *anti,* usually *anti* in DNA and RNA helixes [55], but *syn* purines orientations can be found in Z DNA and non-helical structures such as G-quadruplexes.

The γ angle determines the orientation of the 5'-CH_2O- relative to the sugar ring, and is usually *ap* or +*sc* in experimental structures, due to the *gauche e*ffect between O4' (the furanose ring oxygen) and O5' (Fig. **3**) [56]. Of these, +*sc* is observed in both A and B type nucleic acid helices [55]. Note that both of the preferred *ap* and +*sc* orientations of γ relative to C3' (since γ is defined along the backbone) are *gauche* with respect to O4' [56].

In naturally occurring nucleotides, the value of P for furanose ring pucker is usually around 0-36° (clustering around C3'-*endo*, N-type) or 144-180° (C2'-*endo*, S-type); this restriction is due to classical stereochemical factors involving the natural β sugar configuration (which presents the base and CH_2 both projecting "up" on the same "*endo*" side of the plane of the sugar ring) and the location of both internal and substituent oxygen atoms of the furanose ring system, creating the potential for *gauche* effects on oxygen substituent orientations (as discussed below). In regard to backbone torsional angles other than γ, those torsions involving the phosphorus atoms (angles α and ζ) are also worthy of note: these angles are usually -*sc* in oligonucleotides, due to a *gauche e*ffect involving the phosphate ester (oxygen lone pairs prefer to be *anti* to the opposing oxygen in the O-P-O system, so C-O-P-O prefers to be *gauche*) [55].

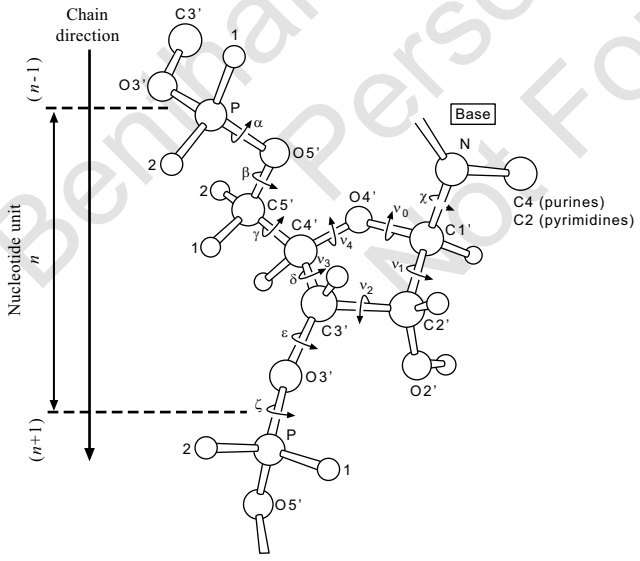

Fig. (5). Naming conventions of the nucleotide subunit and backbone, shown for RNA (2'-OH is present). The backbone torsion angles α, β, γ, δ, ε and ζ are indicated, labeled in the 5' to 3' direction, starting at the 5'-P atom.

4. CLASSICAL STEREOCHEMICAL FACTORS

Classical stereochemical factors can largely explain the observed conformational preferences of nucleosides and nucleotides [50, 51, 55]. The position of the ring oxygen atom (O4') is a major determinant of the conformational preferences of furanoses. Preferred conformations are those involving C2' and/or C3' as the out of plane ring atom(s); since C2' and C3' are opposite O4', which has no substituents, eclipsing interactions are thus minimized. The least favorable conformation is that with O4' *exo,* in which substituents on C2' and C3' are eclipsing and both the bulky β-substituents (C5' and the base) are maximally axial. This results in a higher potential energy barrier to transitions proceeding *via* the west side the pseudorotation cycle, as opposed to the eastern route; these present barriers of about 6 kcal/mole (W) *vs.* 2 kcal/mole (E) relative to the minima at N and S, which are approximately equal for a generic nucleoside, but with a modest preference for S type in 2'-deoxyribonucleotides and N type in ribonucleotides [56, 57].

Classical conformational analysis based primarily on steric interactions predicts that preferred furanose conformations for nucleosides should be around the regions shaded in the cycle (Fig. **4**), *i.e.,* the $^3T_2$ (P = 0°) and $^2T_3$ (P = 180°) twist conformations and the nearby ^3E (C3'-*endo*, P = 18°) or ^2E (C2'-*endo*, P = 162°) envelopes, followed by the $_3E$ (C3'-*exo*, P = 198°) and $_2E$ (C2'-*exo*, P = 342°) envelopes. The latter is somewhat less favorable, since in these, *one* of either the 5'-CH_2OH or the base, respectively, are (pseudo)axial. Due to steric clashes, the west O4' envelope conformation (P= 270°) should be the *least* favorable, as discussed above. These regions correspond very well with experimental observations of nucleic acids and isolated nucleosides [51, 55]. Hence, it is simply the need to stagger rather than eclipse ring substituents, and steric interactions leading to preferred pseudoequatorial rather than pseudoaxial positioning of bulky substituents, that underlie the restriction of nucleosides to the narrow range of N and S types shaded in Fig. (**4**), and explain the relative energy barriers in the E and W of the pseudorotation cycle [56].

Some subtle electronic effects must be invoked, however, to fully explain the conformational preferences of nucleosides and nucleotides. As first pointed out by Olsen in 1982 [57], and as reviewed by Taylor *et al.* [56], the presence of multiple ethylene glycol moieties embedded in DNA and RNA (specifically O4'-C4'-C3'-O3' in both DNA and RNA, and O4'-C1'-C2'-O2' in RNA) leads to opposing *gauche* effects involving O4' and one or both of the 2'-OH (in RNA only) or O3' (in DNA and RNA). An N-type *gauche* orientation of O2' in RNA A-type helices is further stabilized by H-bonding of O2' to the ring O4' of the adjacent nucleoside [55]. Lack of O2' in deoxyribose allows the O4'-C4'-C3'-O3' *gauche* effect to dominate in DNA, leading to the preference of 2'-deoxynucleotides for S-type conformations (P near 180) relative to ribonucleotides, resulting in B type helical preference of DNA [56, 58, 59]. Finally, as mentioned previously, *gauche* effects are also important in restricting certain oligonucleotide backbone torsional angles (α, γ and ζ) and the phosphate moieties to the relatively few observed regions [55]. Various explanations have been offered for the electronic origins of the gauche effect, including one due to

Allinger involving double bond - no bond resonance [60], which is most effective when the torsional angle between the 1,4 electronegative atoms is 90°. Others have invoked quantum mechanical analyses of vicinal interactions between bonding and anti-bonding orbitals [61, 62].

A closely related stereoelectronic effect in sugars is the anomeric effect [46, 57, 63], introduced above, which is very similar to the *gauche* effect for the nucleotide phosphate ester mentioned previously. The oxygen lone pairs in the system LP-O-C-O prefer to be *anti* to the opposing oxygen, so that sugars linked by glycosidic bonds prefer to have the exocyclic OR group axial, in contrast to the expectations of classical stereochemistry, which predicts it should be equatorial because of steric factors. Nucleosides have an analogous O-C-N anomeric effect, in which the lone pairs of the ring oxygen prefer to be anti to the base N1 [35, 59, 64]. This has generally been neglected in nucleic acid modeling, because it exerts an effect that is energetically opposed to the east-west energy barrier to pseudorotation. The preferred anomeric state with a lone pair staggered and opposed (*i.e., anti*) to N1, is in the west of the cycle, whereas the less favored state, with a *carbon* staggered and opposed to N1, is in the east. Jalluri *et al.* [35] proposed that the O-C-N anomeric effect could contribute to the eastern barrier to nucleoside pseudorotation. This investigation also showed that the 1993 vintage of a standard macromolecular force field (Kollman) failed to account for the substantial energy barrier to N-S transition, which had been demonstrated by NMR [65], resulting in the furanose ring converting between N and S forms too easily during dynamics simulations. Based on QM calculations of model compounds, Jalluri *et al.* proposed a simple fix to the existing force field, by the addition of a single new C-O-C-N torsional parameter [35]. In 1999, a more rigorously parameterized version of this approach was implemented by Kollman's group [66], as a modification to the 1995 Cornell *et al.* version of AMBER [67]. Cheatham *et al.* showed that the addition of a new C-O-C-N torsional parameter accounting for the O-C-N anomeric effect, along with other minor changes, led to much more realistic behavior of nucleic acids in MD simulations [66]. One must assume that post-1999 versions of the Kollman force field, therefore, do account for the contribution of the anomeric effect to pseudorotation. We are not sure if other leading macromolecular force fields ever followed their lead in this regard. As to possible mechanisms by which methylation of DNA bases may result in modifications in DNA flexibility and conformational preferences, one possibility that merits future investigation involves the potential impact of base methylation on furanose conformation *via* this O-C-N anomeric effect. Alterations in methylation, or other base modifications, could alter the charge distribution in the base and potentially result in alterations to the degree of polarization of the C1'–N1 bond, and thereby influence pseudorotational barriers in the furanose ring system. A rigorous study of this possibility at the scale of dsDNA would probably require a judicious combination of MM and QM methods to accurately model an effect that depends on rather subtle electronic differences, which MM alone is probably inadequate to address.

CONCLUSION

The remarkable restriction of nucleic acids to relatively few well-defined canonical forms makes them very suitable for study by modeling methods, because their conformational preferences are well understood, and they often represent a more "known" aspect within a study. The considerations reviewed here highlight the need for a critical assessment of fundamental parameters, strengths and weaknesses of computational methods used in nucleic acid simulations. A firm understanding of the known conformational behavior and preferences of biopolymers should be a prerequisite to any modeling study, and it starts at the monomer level. Also, it is important to understand the limitations of the computational methods that are being used. To proceed otherwise, especially as computational methods almost always have inherent limitations, is to risk an outcome that will reflect a lack of background knowledge.

Detection of any large-scale conformation changes in DNA exerted by the modification of nucleic acids bases will require the utmost accuracy in force field methods and parameters, including accurate modeling of higher order torsional effects like *gauche* and anomeric, explicit solvent and counter-ion placement. There are also unique challenges in accurately parameterizing modified bases, which are not recognized by default by many programs in common use. The development and application of hybrid MM/QM methods may be the best path forward. Hopefully this review has outlined some of the considerations that might be of assistance to anyone attempting such a challenge.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Llinàs-Arias, P.; Esteller, M. Epigenetic inactivation of tumour suppressor coding and non-coding genes in human cancer: An update. *Open Biol.,* **2017**, *7*(9), pii: 170152. http://dx.doi.org/10.1098/rsob.170152 PMID: 28931650

[2] Hardwick, J.S.; Lane, A.N.; Brown, T. Epigenetic modifications of cytosine: Biophysical properties, regulation and function in mammalian DNA. *BioEssays,* **2018**, *40*(3), 1700199. http://dx.doi.org/10.1002/bies.201700199 PMID: 29369386

[3] Varriale, A. DNA methylation, epigenetics, and evolution in vertebrates: Facts and challenges. *Int. J. Dev. Biol.,* **2014**, 475981. http://dx.doi.org/10.1155/2014/475981

[4] Kumar, S.; Cheng, X.; Klimasauskas, S.; Mi, S.; Posfai, J.; Roberts, R.J.; Wilson, G.G. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.,* **1994**, *22*(1), 1-10. http://dx.doi.org/10.1093/nar/22.1.1 PMID: 8127644

[5] Ferreira, Humberto; Manel, E CpG Islands in cancer: heads, tails, and sides. *Methods Enzymol.,* **2018**, *1766*(CpG Islands), 49-80.

[6] López, V.; Fernández, A.F.; Fraga, M.F. The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. *Ageing Res. Rev.,* **2017**, *37*, 28-38. http://dx.doi.org/10.1016/j.arr.2017.05.002 PMID: 28499883

[7] Eleftheriou, M.; Pascual, A.J.; Wheldon, L.M.; Perry, C.; Abakir, A.; Arora, A.; Johnson, A.D.; Auer, D.T.; Ellis, I.O.; Madhusudan, S.; Ruzov, A. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical Epigenetics,* **2015**, *7, 88.*

[8] Jeltsch, A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem,* **2002**, *3*(4), 274-293. http://dx.doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S PMID: 11933228

[9] Gruber, D.R.; Toner, J.J.; Miears, H.L.; Shernyukov, A.V.; Kiryutin, A.S.; Lomzov, A.A.; Endutkin, A.V.; Grin, I.R.; Petrova, D.V.; Kupryushkin, M.S.; Yurkovskaya, A.V.; Johnson, E.C.; Okon, M.; Bagryanskaya, E.G.; Zharkov, D.O.; Smirnov, S.L. Oxidative damage to epigenetically methylated sites affects DNA stability, dynamics and enzymatic demethylation. *Nucleic Acids Res.,* **2018**, *46*(20), 10827-10839. http://dx.doi.org/10.1093/nar/gky893 PMID: 30289469

[10] Lee, J.Y.; Lee, T.H. Effects of DNA methylation on the structure of nucleosomes. *J. Am. Chem. Soc.,* **2012**, *134*(1), 173-175. http://dx.doi.org/10.1021/ja210273w PMID: 22148575

[11] Lapinska, K.; Faria, G.; McGonagle, S.; Macumber, K.M.; Heerboth, S.; Sarkar, S. Cancer progenitor cells: The result of an epigenetic event? *Anticancer Res.,* **2018**, *38*(1), 1-6. PMID: 29277749

[12] Zhou, Z.; Sun, B.; Li, X.; Zhu, C. DNA methylation landscapes in the pathogenesis of type 2 diabetes mellitus. *Nutr. Metab. (Lond.),* **2018**, *15*(47), 47. http://dx.doi.org/10.1186/s12986-018-0283-x PMID: 29988495

[13] Potaczek, D.P.; Harb, H.; Michel, S.; Alhamwe, B.A.; Renz, H.; Tost, J. Epigenetics and allergy: From basic mechanisms to clinical applications. *Epigenomics,* **2017**, *9*(4), 539-571. http://dx.doi.org/10.2217/epi-2016-0162 PMID: 28322581

[14] Liang, L.; Willis-Owen, S.A.G.; Laprise, C.; Wong, K.C.C.; Davies, G.A.; Hudson, T.J.; Binia, A.; Hopkin, J.M.; Yang, I.V.; Grundberg, E.; Busche, S.; Hudson, M.; Rönnblom, L.; Pastinen, T.M.; Schwartz, D.A.; Lathrop, G.M.; Moffatt, M.F.; Cookson, W.O.C.M. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature,* **2015**, *520*(7549), 670-674. http://dx.doi.org/10.1038/nature14125 PMID: 25707804

[15] Duan, Y.T.; Sangani, C.B.; Liu, W.; Soni, K.V.; Yao, Y. New promises to cure cancer and other genetic diseases/disorders: Epi-drugs through epigenetics. *Curr. Top. Med. Chem.,* **2019**, *19*(12), 972-994. http://dx.doi.org/10.2174/1568026619666190603094439 PMID: 31161992

[16] Yang, K.; Park, D.; Tretyakova, N.Y.; Greenberg, M.M. Histone tails decrease N7-methyl-2′-deoxyguanosine depurination and yield DNA-protein cross-links in nucleosome core particles and cells. *Proc. Natl. Acad. Sci. USA,* **2018**, *115*(48), E11212-E11220. http://dx.doi.org/10.1073/pnas.1813338115 PMID: 30429328

[17] Singer, B.; Hang, B. Nucleic acid sequence and repair: role of adduct, neighbor bases and enzyme specificity. *Carcinogenesis,* **2000**, *21*(6), 1071-1078. http://dx.doi.org/10.1093/carcin/21.6.1071 PMID: 10836993

[18] Tan, X.; Suzuki, N.; Grollman, A.P.; Shibutani, S. Mutagenic events in Escherichia coli and mammalian cells generated in response to acetylaminofluorene-derived DNA adducts positioned in the Nar I restriction enzyme site. *Biochemistry,* **2002**, *41*(48), 14255-14262. http://dx.doi.org/10.1021/bi0202878 PMID: 12450390

[19] Bestor, T.; Edwards, J.R.; Ju, J.; Li, X. Universal methylation profiling methods. *IPN WO,* **2010**, 2010/011312, A9.

[20] Rollins, R.A.; Haghighi, F.; Edwards, J.R.; Das, R.; Zhang, M.Q.; Ju, J.; Bestor, T.H. Large-scale structure of genomic methylation patterns. *Genome Res.,* **2006**, *16*(2), 157-163. http://dx.doi.org/10.1101/gr.4362006 PMID: 16365381

[21] Yan, S.; Shapiro, R.; Geacintov, N.E.; Broyde, S. Stereochemical, structural, and thermodynamic origins of stability differences between stereoisomeric benzo[a]pyrene diol epoxide deoxyadenosine adducts in a DNA mutational hot spot sequence. *J. Am. Chem. Soc.,* **2001**, *123*(29), 7054-7066. http://dx.doi.org/10.1021/ja0043035 PMID: 11459484

[22] Garner, R.C.; Tierney, B.; Phillips, D.H. A comparison of ^{32}P-postlabelling and immunological methods to examine human lung DNA for benzo[a]pyrene adducts. *IARC Sci. Publ.,* **1988**, *89*(89), 196-200. PMID: 3143669

[23] van Schooten, F.J.; Hillebrand, M.J.; Scherer, E.; den Engelse, L.; Kriek, E. Immunocytochemical visualization of DNA adducts in mouse tissues and human white blood cells following treatment with benzo[a]pyrene or its diol epoxide. A quantitative approach. *Carcinogenesis,* **1991**, *12*(3), 427-433. http://dx.doi.org/10.1093/carcin/12.3.427 PMID: 1901249

[24] Baan, R.A.; Steenwinkel, M.J.; van Asten, S.; Roggeband, R.; van Delft, J.H. The use of benzo[a]pyrene diolepoxide-modified DNA standards for adduct quantification in ^{32}P-postlabelling to assess exposure to polycyclic aromatic hydrocarbons: application in a biomonitoring study. *Mutat. Res.,* **1997**, *378*(1-2), 41-50. http://dx.doi.org/10.1016/S0027-5107(97)00096-1 PMID: 9288884

[25] Guo, l.; Jiang, X.; Tian, H.-Y.; Yao, S.J.; Li, B.-Y.; Zhang, R.J.; Zhang, S.-S.; Sun, X. Detection of BPDE-DNA adducts in human umbilical cord blood by LC-MS/MS analysis. *Yao Wu Shi Pin Fen Xi,* **2019**, *27*, 518-525.

[26] Glick, J.; Xiong, W.; Lin, Y.; Noronha, A.M.; Wilds, C.J.; Vouros, P. The influence of cytosine methylation on the chemoselectivity of benzo[a]pyrene diol epoxide-oligonucleotide adducts determined using nanoLC/MS/MS. *J. Mass Spectrom.,* **2009**, *44*(8), 1241-1248. http://dx.doi.org/10.1002/jms.1605 PMID: 19536795

[27] Cheng, S.C.; Hilton, B.D.; Roman, J.M.; Dipple, A. DNA adducts from carcinogenic and noncarcinogenic enantiomers of benzo[a]pyrene dihydrodiol epoxide. *Chem. Res. Toxicol.,* **1989**, *2*(5), 334-340. http://dx.doi.org/10.1021/tx00011a011 PMID: 2519824

[28] Denissenko, M.F.; Chen, J.X.; Tang, M.S.; Pfeifer, G.P. Cytosine methylation determines hot spots of DNA damage in the human P53 gene. *Proc. Natl. Acad. Sci. USA,* **1997**, *94*(8), 3893-3898. http://dx.doi.org/10.1073/pnas.94.8.3893 PMID: 9108075

[29] Kolbanovskiy, A.; Kuzmin, V.; Shastry, A.; Kolbanovskaya, M.; Chen, D.; Chang, M.; Bolton, J.L.; Geacintov, N.E. Base selectivity and effects of sequence and DNA secondary structure on the formation of covalent adducts derived from the equine estrogen metabolite 4-hydroxyequilenin. *Chem. Res. Toxicol.,* **2005**, *18*(11), 1737-1747. http://dx.doi.org/10.1021/tx050190x PMID: 16300383

[30] Cheatham, T.E., III; Young, M.A. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers,* **2000-2001**, *56*(4), 232-256. http://dx.doi.org/10.1002/1097-0282(2000)56:4<232::AID-BIP10037>3.0.CO;2-H PMID: 11754338

[31] Naômé, A.; Schyman, P.; Laaksonen, A.; Vercauteren, D.P. Molecular dynamics simulation of 8-oxoguanine containing DNA fragments reveals altered hydration and ion binding patterns. *J. Phys. Chem. B,* **2010**, *114*(14), 4789-4801. http://dx.doi.org/10.1021/jp1000539 PMID: 20307074

[32] Fischer, N.M.; Polêto, M.D.; Steuer, J.; van der Spoel, D. Influence of Na^+ and Mg^{2+} ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res.,* **2018**, *46*(10), 4872-4882. http://dx.doi.org/10.1093/nar/gky221 PMID: 29718375

[33] Cheatham, T.E., III; Kollman, P.A. Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.,* **1997**, *119*, 4805-4825. http://dx.doi.org/10.1021/ja963641w

[34] Cheatham, T.E., III Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.,* **2004**, *14*(3), 360-367. http://dx.doi.org/10.1016/j.sbi.2004.05.001 PMID: 15193317

[35] Jalluri, R.; Yuh, Y.H.; Taylor, E.W. The O-C-N anomeric effect in nucleosides: a major factor underlying the experimentally observed eastern barrier to pseudorotation. *ACS Symposium Series.* **1993**, *Vol. 539,* pp. 277-293. http://dx.doi.org/10.1021/bk-1993-0539.ch015

[36] Zacharias, M. Conformational analysis of DNA-trinucleotide-hairpin-loop structures using a continuum solvent model. *Biophys. J.,* **2001**, *80*(5), 2350-2363. http://dx.doi.org/10.1016/S0006-3495(01)76205-4 PMID: 11325735

[37] Gilson, M.K.; Davis, M.E.; Luty, B.A.; McCammon, J.A. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.,* **1993**, *97*, 3591-3600. http://dx.doi.org/10.1021/j100116a025

[38] Pérez, A.; Luque, F.J.; Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.,* **2012**, *45*(2), 196-205. http://dx.doi.org/10.1021/ar2001217 PMID: 21830782

[39] Kumari, I.; Sandhu, P.; Ahmed, M.; Akhter, Y. Molecular dynamics simulations, challenges and opportunities: A biologist's prospective. *Curr. Protein Pept. Sci.,* **2017**, *18*(11), 1163-1179. http://dx.doi.org/10.2174/1389203718666170622074741 PMID: 28637405

[40] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.,* **1993**, *98*, 10089-10092. http://dx.doi.org/10.1063/1.464397

[41] Halgren, T.A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.,* **1999**, *20*, 720-729. http://dx.doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X

[42] Halgren, T.A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.,* **1999**, *20*, 730-748. http://dx.doi.org/10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2-T

[43] Nester, K.; Gaweda, K.; Plazinski, W. A GROMOS force field for furanose-based carbohydrates. *J. Chem. Theory Comput.,* **2019**, *15*(2), 1168-1186. http://dx.doi.org/10.1021/acs.jctc.8b00838 PMID: 30609362

[44] Wavefunction, Inc.; Irvine, CA 92612, USA.

[45] Andrews, C.W.; Fraser-Reid, B.; Bowen, J.P. An *ab initio* study (6-31G*) of transition states in glycoside hydrolysis based on axial and equatorial 2-methoxytetrahydropyrans. *J. Am. Chem. Soc.,* **1991**, *113*, 8293-8298. http://dx.doi.org/10.1021/ja00022a015

[46] Wiberg, K.B.; Bailey, W.F.; Lambert, K.M.; Stempel, Z.D. The anomeric effect: It's complicated. *J. Org. Chem.,* **2018**, *83*(9), 5242-5255. http://dx.doi.org/10.1021/acs.joc.8b00707 PMID: 29620891

[47] Aranda, J.; Zinovjev, K.; Swiderek, K.; Roca, M.; Tuñón, I. Unraveling the reaction mechanism of enzymatic c5-cytosine methylation of DNA. A combined molecular dynamics and QM/MM study of wild type and Gln119 variant. *ACS Catal.,* **2016**, *6*, 3262-3276. http://dx.doi.org/10.1021/acscatal.6b00394

[48] Harris, S.A. Modelling the biomechanical properties of DNA using computer simulation. *Philos. Trans.- Royal Soc., Math. Phys. Eng. Sci.,* **2006**, *364*(1849), 3319-3334. http://dx.doi.org/10.1098/rsta.2006.1906 PMID: 17090462

[49] Young, M.A.; Ravishanker, G.; Beveridge, D.L. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J.,* **1997**, *73*(5), 2313-2336. http://dx.doi.org/10.1016/S0006-3495(97)78263-8 PMID: 9370428

[50] Sundaralingam, M. Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers,* **1969**, *7*, 821-860. http://dx.doi.org/10.1002/bip.1969.360070602

[51] Sundaralingam, M. The stereochemical concepts of nucleic acids. *Conformation in Biology*; Srinivasan, R.; Sarma, R.H., Eds.; Adenine Press: NY, **1983**, pp. 191-225.

[52] Juaristi, E. The attractive and repulsive gauche effects. *J. Chem. Educ.,* **1979**, *56*, 438-441. http://dx.doi.org/10.1021/ed056p438

[53] Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.,* **1972**, *94*(23), 8205-8212. http://dx.doi.org/10.1021/ja00778a043 PMID: 5079964

[54] Westhof, E.; Sundaralingam, M. A Method for the analysis of puckering disorder in five-membered rings: The relative mobilities of furanose and proline rings and their effects on polynucleotide and polypeptide backbone flexibility. *J. Am. Chem. Soc.,* **1983**, *105*, 970-976. http://dx.doi.org/10.1021/ja00342a054

[55] Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, N.Y., **1984**. http://dx.doi.org/10.1007/978-1-4612-5190-3

[56] Taylor, E.W.; Van Roey, P.; Schinazi, R.F.; Chu, C.K. A stereochemical rationale for the activity of anti-HIV nucleosides. *Antivir. Chem. Chemother.,* **1990**, *1*, 163-173. http://dx.doi.org/10.1177/095632029000100301

[57] Olson, W.K. How flexible is the furanose ring? II. An updated potential energy estimate. *J. Am. Chem. Soc.,* **1982**, *104*, 278-286. http://dx.doi.org/10.1021/ja00365a050

[58] Pearlman, D.A.; Kim, S-H. Conformational studies of nucleic acids. II. The conformational energetics of commonly occurring nucleosides. *J. Biomol. Struct. Dyn.,* **1985**, *3*(1), 99-125. http://dx.doi.org/10.1080/07391102.1985.10508401 PMID: 3917020

[59] Plavec, J.; Tong, W.; Chattopadhyaya, J. How do the gauche and anomeric effects drive the pseudorotational equilibrium of the pentofuranose moiety of nucleosides? *J. Am. Chem. Soc.,* **1993**, *115*, 9734-9746. http://dx.doi.org/10.1021/ja00074a046

[60] Allinger, N.L.; Hindman, D.; Hornig, H. Conformational analysis. 125. The importance of twofold barriers in saturated molecules. *J. Am. Chem. Soc.,* **1977**, *99*, 3282-3284. http://dx.doi.org/10.1021/ja00452a014

[61] Radom, L.; Hehre, W.J.; Pople, J.A. Molecular orbital theory of the electronic structure of organic compounds. XIII. Fourier component analysis of internal rotation potential functions in saturated molecules. *J. Am. Chem. Soc.,* **1972**, *94*, 2371-2381. http://dx.doi.org/10.1021/ja00762a030

[62] Brunck, T.K.; Weinhold, F. Quantum-mechanical studies on the origin of barriers to internal rotation about single bonds. *J. Am. Chem. Soc.,* **1979**, *1979*(101), 1700-1709. http://dx.doi.org/10.1021/ja00501a009

[63] Kirby, A.J. *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*; Springer-Verlag: Berlin, **1983**. http://dx.doi.org/10.1007/978-3-642-68676-4

[64] Koole, L.H.; Buck, H.M.; Nyilas, A.; Chattopadhyaya, J. Structural properties of modified deoxyadenosine structures in solution. Impact of the gauche and anomeric effects on the furanose conformation. *Can. J. Chem.,* **1987**, *65*, 2089-2094. http://dx.doi.org/10.1139/v87-346

[65] Röder, O.; Lüdemann, H.; Von Goldammer, E. Determination of the activation energy for pseudorotation of the furanose ring in nucleosides by ^{13}C nuclear-magnetic-resonance relaxation. *Eur. J. Biochem.,* **1975**, *53*(2), 517-524. http://dx.doi.org/10.1111/j.1432-1033.1975.tb04094.x PMID: 1140200

[66] Cheatham, T.E., III; Cieplak, P.; Kollman, P.A. A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.,* **1999**, *16*(4), 845-862. http://dx.doi.org/10.1080/07391102.1999.10508297 PMID: 10217454

[67] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M., Jr; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.,* **1995**, *117*, 5179-5197. http://dx.doi.org/10.1021/ja00124a002