Computational Approaches for Investigating Base Flipping in Oligonucleotides†

U. Deva Priyakumar and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201

Received April 29, 2005

Contents

1. Introduction

1.1. Overview of Base Flipping in Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA)

Protein−DNA interactions are central for transcription, replication, and various DNA repair mechanisms.[1−5] Some proteins exert their effects by simple binding to DNA or RNA without much structural deformation of the latter, while in other cases binding involves extreme structural distortion of the oligonucleotide. Among these are proteins capable of opening specific base pairs to perform chemical reactions on the target base. These include enzymes such as methyltransferases, polymerases, nucleases, glycosylases, integrases, and recombinases.[6−10]

When a chemical reaction is performed on a DNA base or bases [referred to as the target base(s)] by a protein, those bases have to be accessible to the protein for the desired chemistry to take place. Originally, the accessibility problem was thought to be solved via distortion of the double helix by bending and kinking, thereby exposing the base to the protein. However, in 1994, the ternary structure of the complex of the methyltransferase from *Hha*I (M.*Hha*I) with a modified duplex 13-mer DNA containing methylated 5-fluorocytosine at the target site and S-adenosyl-L-homocysteine (SAH) was reported by Klimasauskas et al.[11] The DNA bound to the enzyme exhibits an unusual conformation where the target base of the DNA is completely swung out of its Watson−Crick (WC) base-paired helical position and bound in the catalytic pocket of the enzyme (Figure 1). This type of structural distortion of the DNA enables the catalytic enzyme to access the specific base and perform chemical reactions on it. For example, M.*Hha*I catalyzes the transfer of methyl group from S-adenosyl-L-methionine (SAM) to the target base cytosine, the mechanism of which has been studied extensively (Figure 2).[12−15] Over the years, numerous crystal structures of protein−DNA complexes where base flipping occurs have been reported, including several methyltransferases (M.*Hha*I,[11,16] M.*Hae*III,[17] and M.*Taq*I[18,19]), glycosylases[20,21] (T4 endonuclease V,[22] human UDG,[23−25] *Escherichia coli* MUG,[26] human AAG,[27] *E. coli* AlkA,[28] and bOGG1[29]) and endonucleases (*E. coli* endonuclease IV[30] and HAP1[31]). Clearly, base flipping, as it is commonly known, is a phenomenon important for the biological function of both DNA and RNA.[6,7,10,32]

1.2. Scope of the Review

The present manuscript will give an overview of the phenomenon of base flipping, with emphasis on the atomic details of the structural and energetic events that dictate its occurrence as obtained via computational approaches. Information presented will include experimental data available on base flipping in DNA alone and in the presence of proteins that is relevant to the interpretation of the computational work. This will be followed by an overview of computational approaches used to study base flipping, including a critical evaluation of those approaches. The discussion will also

† This paper is dedicated to the memory of Dr. Kate DeTurck, a wonderful person, scientist, and veterinarian that enjoyed the flip side of life.

* To whom correspondence should be addressed. Tel: 410-706-7442. Fax: 410-706-5017. E-mail: amackere@rx.umaryland.edu.

10.1021/cr040475z CCC: $59.00
Published on Web 12/23/2005

Professor Alexander D. MacKerell, Jr. received an Associate's Degree in Biology from Gloucester County College in 1979, a Bachelor's of Science in Chemistry from the University of Hawaii in 1981, and a Ph.D. in Biochemistry from Rutgers University in 1985. Postdoctoral studies were performed at the Karolinska Institutet, Stockholm, Sweden, from 1986 to 1988 and at Harvard University from 1988 to 1992. Following one year as a Visiting Professor at Swarthmore College, Prof. MacKerell moved to the University of Maryland, School of Pharmacy, where he has remained. His research interests include the application and development of theoretical methods to systems of chemical, biological, and pharmacological interest. His efforts have involved the development of empirical force field parameters associated with the CHARMM program, structure–function studies of nucleic acids and proteins, and computer-aided drug design (CADD) studies in the areas of opioids, immunosuppressants, anticancer agents, and antibiotics. He is also currently director of the CADD Center at the University of Maryland.

U. Deva Priyakumar received his B.Sc. in Chemistry from St. Joseph's College (University of Madras, India) in 1997. Later, he obtained M.Sc. (1999), M.Phil. (2000), and Ph.D. (2003) degrees in Chemistry from Pondicherry University, India. During his Ph.D., he worked on the application of various quantum mechanical approaches in computing structural, energetic, magnetic, and spectroscopic properties and reactivity parameters of organic and organometallic compounds under the direction of Dr. G. Narahari Sastry. He joined the group of Alex MacKerell at the University of Maryland Baltimore as a postdoctoral fellow in 2004. Currently, he is involved in elucidating base flipping mechanisms and DNA and RNA simulations and identifying methyltransferace inhibitors using computer-aided drug design. His research interest in general is the application of computational methods for understanding why chemical molecules and biological macromolecules behave the way that they do.

include how available experimental data can be used to validate the computed methodologies. Enormous amounts of data are available on the structural, mechanistic, and kinetic properties of various base-flipping enzymes and has been reviewed elsewhere.[3,4,9,20,21,34−39] Emphasis in the present review will be on flipping associated with the enzyme M.*Hha*I, due to the significant volume of data on that protein

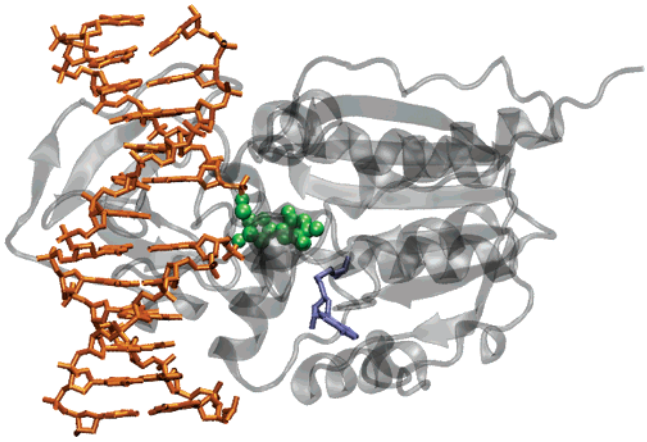

Figure 1. Crystal structure of the ternary complex of M.*Hha*I, DNA, and SAH (PDB ID: 1MHT) generated using VMD.[33] The target base cytosine (CPK representation in green) of the DNA is flipped out of the duplex structure and is bound to the catalytic site of the enzyme (ribbon representation). The coenzyme, SAH, is depicted using thick bonds in blue.

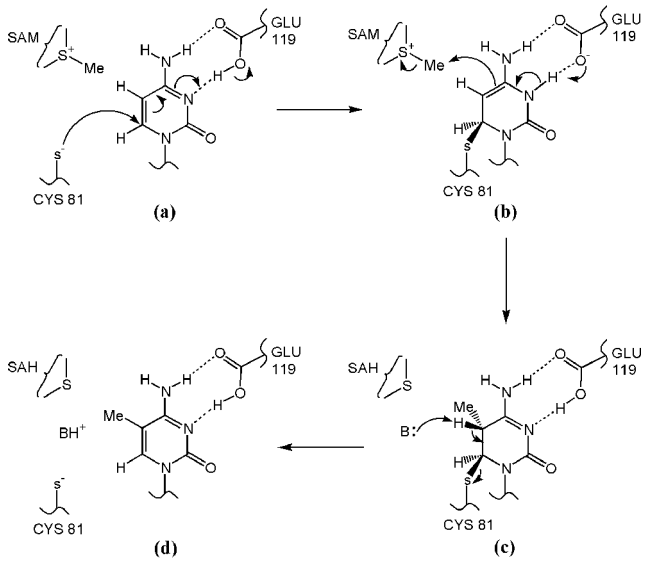

Figure 2. Schematic representation of the methyl transfer mechanism catalyzed by M.*Hha*I. The target base, cytosine (a), forms a covalent complex (b) with the protein followed by a methyl transfer from SAM (c) and finally dissociation of cytosine from the protein to yield 5-methylcytosine (d).

as well as the familiarity of the authors with that system. Upon completion of this review, it is hoped that the reader will have a better understanding of the power of computational methods in elucidating details of this biologically essential structural change and how those approaches may be applied to base flipping as well as other structural perturbations in oligonucleotides.

2. Experimental Background

2.1. Spontaneous Base Opening in Nucleic Acids

Oligonucleotides are flexible biomolecules that undergo a variety of conformational changes essential for their biological function; one such motion is base opening or flipping. The most common experimental method to measure base opening is the exchange of the imino protons in G, T, or U bases monitored by nuclear magnetic resonance (NMR) spectroscopy.[40−45] This approach has been extensively used

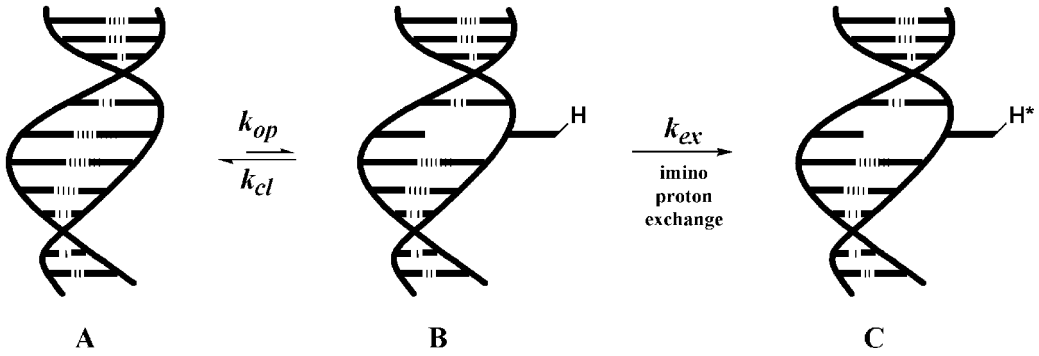

Figure 3. Schematic representation of the base opening−closing and imino proton exchange processes. The canonical form of DNA or RNA (A) in its minimum energy conformation exists in equilibrium with the base-opened state (B) with the equilibrium favored toward the left-hand side. The imino proton of B is exchanged by a solvent molecule (not shown) to yield C (the proton from the solvent is denoted as H*). The opening base need not flip out completely for the proton exchange to take place (see Figure 4). k_{op}, k_{cl}, and k_{ex} are the opening rate, closing rate, and rate of the exchange process, respectively. k_{ex} is assumed to be equal to k_{op} as base opening is the rate-limiting step under proper conditions.

to evaluate the base-opening rates in DNA and RNA duplexes. Other methods used to experimentally study base opening include trapping experiments, as discussed below, as well as isotope exchange studies or via ultraviolet, infrared, or Raman spectroscopic techniques, although the latter have only met with limited success.

NMR may be considered the technique of choice as it allows the proton exchange to be unequivocally assigned to a specific base pair. Once the base pair opens from its original position in the duplex, the imino proton becomes accessible to the solvent environment and proton exchange occurs (Figure 3). The calculated imino proton exchange rate may be assumed to be equal to the base-opening rate provided that the exchange is fast enough that it occurs at every base-opening event. This condition is achieved using proton acceptors (i.e., base-catalyzed exchange), such as ammonia, and extrapolating the experimentally measured exchange rates to infinite acceptor concentration.

Base pair opening has been studied via imino proton exchange in a variety of nucleic acid structures including B-DNA,[42] Z-DNA,[46] RNA,[47] and tRNA.[48] The base pair lifetimes are typically on the order of milliseconds. Opening events, at room temperature, are thought to involve a single base pair with minimal perturbation of the remainder of the local oligonucleotide structure. Lifetimes of the open state of base pairs range from 10 to a few hundreds of nanoseconds. Generally, the lifetimes for A:T and G:C WC pairs are observed to be 1−5 and 10−50 ms, respectively. While the base opening is primarily dictated by the strength of base pair interactions, the role of sequence context on opening rates has been shown.[44,45,49−51] The variation in the rates is not large in most of the cases; however, interesting trends in the exchange rates have been observed. AT base pairs in A-tracts have longer lifetimes (≥100 ms) as compared to other sequences.[50] In contrast, GC base pair lifetimes in G-tracts are much shorter than that observed normally.[49] The GC base pair lifetimes in RNA range from 40 to 50 ms and in general are longer as compared to that in DNA, whereas for the AU base pairs lifetimes are less than 1 ms.[47]

These experimental studies have been extremely helpful in understanding base pair opening in nucleic acids alone. However, the rate constants obtained for the DNA cannot be strictly compared to the flipping in the presence of enzymes for the following reasons. First, in enzymatic base flipping, the target base undergoes a significant rotation to assume an extrahelical position that is often almost 180° out of the DNA duplex. With imino proton exchange studies, the base only has to open enough for the imino protons to exchange with the acceptor base in the surrounding solvent. This requires the technique to rely on a two state model assumption, where the base is considered open or closed irrespective of the extent of opening. As will be discussed below, theoretical calculations suggest that an opening angle of only ±30° is necessary for the imino proton exchange to take place.[52−54] Figure 4 depicts the structure of duplex DNA, the fully flipped state, and some representative conformations for the intermediate structures involved in base flipping. As described, conformers B and C can undergo proton exchange similar to the fully flipped state (F) or other conformations with the base open to a similar or larger extent (for example, D and E). However, states B and C are not fully base-flipped states with respect to the flipping in the presence of enzymes. Therefore, all of the conformations that undergo exchange cannot be considered as base-flipped states. Second, base-opening rates assessed using imino proton exchange experiments correspond to the base pair and not to a specific base. For example, a G:C opening rate corresponds to the opening of both G and C. It has been shown using computational experiments that the imino proton of G in its helical conformation is accessible to the solvent for exchange when its WC base-paired counterpart cytosine is flipped out.[52] Hence, it is not possible to experimentally measure the kinetics of the opening of individual bases. However, in the case of enzymatic base flipping, the kinetic measurements correspond to only the target base, which flips out and binds to the catalytic pocket of the enzyme, whereas its WC partner remains in its original position in the DNA duplex. Thus, the rate constants measured from imino proton exchange may not necessarily be appropriate for the interpretation of enzymatic flipping studies.

2.2. Trapping of Flipped Base by Macrocycle Host Molecules

Methods to overcome the limitation discussed above with imino proton exchange measurements are approaches where the base is trapped in a fully flipped state. The base must be fully extruded from the helix for a macrocycle (host molecule) to bind to the flipped base. Such a scenario may be considered more analogous to that occurring in enzymatic flipping studies. Accordingly, rate constants from such studies, as discussed in the following paragraphs, may be considered to be more appropriate for the interpretation of

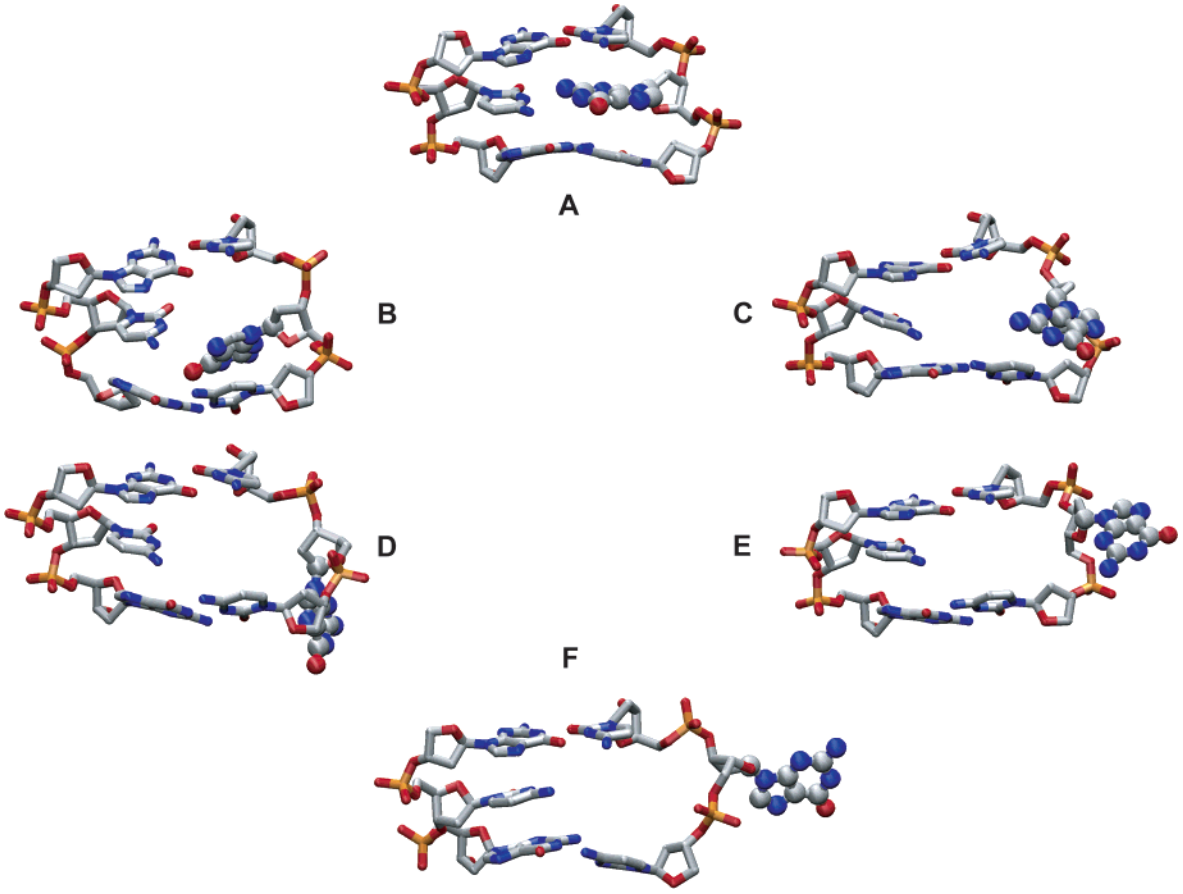

Figure 4. Representative conformations of the duplex DNA (A, opening angle ~0°), the fully flipped state (F, opening angle ~180°), and representative intermediates (B, C, D, and E) during the base-flipping process with opening angles of 60, −60, 120, and −120°, respectively. For the imino proton exchange to occur, the base need not flip out completely; instead, an opening of approximately ±30° is suggested to be enough for the exchange to occur.

protein-induced base flipping. However, the number of studies involving the trapping of a flipped base is limited.

β-Cyclodextrin has been shown to form stable host−guest complexes with adenine.[55,56] While γ-cyclodextrin forms weak inclusion complexes, the guest cavity in α-cyclodextrin is too small to accommodate the large purine bases.[57] Spies and Schowen have used β-cyclodextrin to trap the flipped base in DNA by taking advantage of the fact that β-cyclodextrin binds strongly to purine bases forming stable host−guest complexes.[58] The effective binding of cyclodextrin to the flipped base is assessed by monitoring the melting point depression of the nucleic acid. When the macrocycle is bound to the oligonucleotide, DNA undergoes melting below the temperature at which it normally melts. It also has been shown via spectroscopic methods that the cyclodextrin molecule does not bind to oligonucleotides whose duplex form is intact, confirming that base flipping did occur prior to binding.[55] However, information about which base undergoes flipping and specificity of binding is not available. Importantly, the macrocycle does not induce base flipping. The rate constant for β-cyclodextrin binding to the flipped base in both DNA and RNA has been shown to be $(3.5 \pm 0.5) \times 10^{-3}$ s^{-1}. This rate is approximately 4−5 orders of magnitude slower than that measured by imino proton exchange, emphasizing the need to consider the appropriate experimental method being used to obtain rate data when investigating base flipping in the presence of proteins.

Another chemical-trapping agent, bisacridine macrocycle (BisA), was found to induce base flipping in DNA containing mismatches.[59] Previously, it was shown that the BisA molecule binds in two different modes to the abasic and unoccupied sites in DNA.[60,61] The structure of one of these complexes was characterized using NMR and molecular-modeling techniques. The structural features reveal that it interacts by the so-called thread-binding mode with the base adjacent to the unoccupied site sandwiched into the cavity of BisA (Figure 5). In other words, one of the acridine moieties of BisA resides in the abasic site and the other heterocyclic structure is flanked by the next two base pairs. Unlike the cyclodextrin complexes, the melting temperature of the 1:1 complex of DNA and BisA was observed to be higher as compared to that of the DNA by itself. Thermal denaturation experiments showed that BisA recognizes a TT, TC, or TG mismatch in a 17-mer duplex and binds to the oligonucleotide resulting in an increase in the melting temperature of the DNA of up to 7.1 °C, whereas for a TA match the melting temperature remains unaltered in the presence of BisA. They have proposed that one of the mismatched bases undergoes flipping assisted by this macrocycle, which later binds to the base adjacent to the unoccupied site (Figure 5).

2.3. Enzymatic Base Flipping

Base flipping as a mechanism involved in the epigenetic control of gene expression and its role in DNA repair involves interactions with a variety of proteins.[6−10] As mentioned above, enzymes that employ a base-flipping mechanism include methyltransferases, glycosylases, and

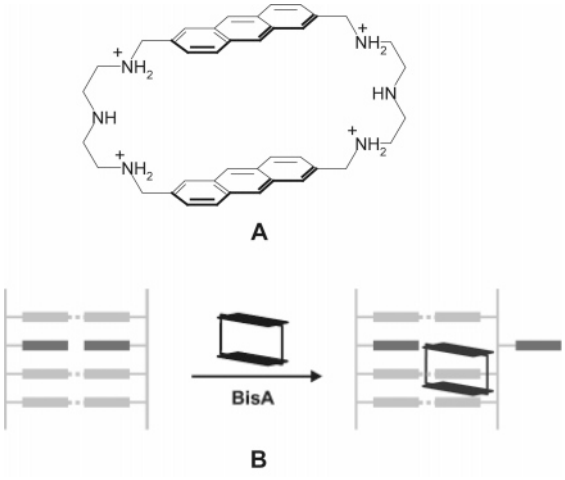

Figure 5. Structure of the Bisacridine macrocycle (A) and the schematic representation of the induced base flipping by BisA and its binding mode to the unoccupied site (B). The mismatched base pair is denoted by darker blocks while the matching base pairs are designated by gray rectangular blocks. BisA appears to flip open one of the mismatches and binds to the unoccupied site sandwiching the adjacent base. Reprinted with permission from ref 59. Copyright 2003 Wiley-VCH.

endonucleases. An important aspect of the action of these enzymes is their ability to act on particular sites on DNA; for example, methyltransferases involved in bacterial restriction systems target specific DNA sequences while DNA repair enzymes have to identify the site of damage.[3,5,9,34,37,62] Accordingly, when investigating flipping in the presence of enzymes, issues associated with the flipping event itself, specificity, and their impact in the mechanism of the enzyme must be considered.

2.3.1. Kinetic Considerations

Enzymes that include base flipping in their reaction mechanism, in general, must first bind to the DNA in a site-specific fashion, followed by chemical catalysis and subsequent dissociation of the oligonucleotide from the protein. Kinetic studies by Wu and Santi revealed that methylation by M.*Hha*I primarily follows an ordered bi bi mechanism, where the DNA binds to the protein, followed by binding of the coenzyme, SAM.[63] Following the methyl transfer from SAM to the target C base, the coenzyme, now S-adenosylhomocysteine, dissociates followed by dissociation of the DNA from the enzyme. Overall, the k_{cat} value for the reaction was observed to be 0.02 s^{-1} and the rate-limiting step could not be assigned based on this study, although the rate-determining step is not the formation of the DNA−enzyme complex. The crystal structure of the binary complex of M.*Hha*I and SAM[64] indicates that a binary complex involving the coenzyme can indeed be formed, which is not a part of the mechanism proposed by Wu and Santi. A recent study by Lindstrom et al. found that the enzyme−SAM complex is formed but exhibits a 50-fold decrease in the affinity as compared to the DNA−M.*Hha*I−SAM ternary complex.[65] Detailed kinetic studies have shown that that rate-limiting step for this reaction could be product release.[13,63,65,66] Kinetic studies on methyltransferases including M.*Hha*I, M.*Mva*I, and *Eco*RI show that the k_{cat} value ranges from 10^{-3} to 10^{-1} s.[63,67,68] Thus, in the case of methyltransferases, the reaction rate is relatively slow, on the order of seconds.

Considering the reaction rate in the presence of the enzyme allows for the question of the role of the protein in base flipping to be addressed; is it an active or passive role? Simply put, does the protein actively facilitate the flipping of the base out of the double helix or does the enzyme wait passively until the base spontaneously flips out of the helix, following which it binds to the now-exposed base? Analysis of the rate constants discussed above can address this question. If the rate constants for flipping based on imino proton exchange are used as the flipping rate in DNA alone (i.e., millisecond time scale), as was initially assumed as that was the only data available, and these are compared to k_{cat} for the enzymatic reactions (second time scale), the fact that the spontaneous flipping rate is 3 orders of magnitude faster than k_{cat} strongly suggests a passive mechanism. However, when the more recent data on flipping rates based on trapping experiments are considered (100−1000 s time scale), an active role of the enzyme in flipping is certainly possible, as the spontaneous flipping rate is an order of magnitude or more slower than the catalytic rate and slower than the rate of base flipping, which is around 195 s^{-1} in the case of *Eco*RI.[69] In addition, as stated above, the rate-determining step in the M.*Hha*I reaction mechanism occurs after methyl transfer, supporting, although not proving, that flipping of the base is being facilitated by the protein such that it is not rate limiting. Thus, consideration of the rate data to use for spontaneous base flipping in DNA alone can lead to significant differences in conclusions concerning an active vs a passive role of base-flipping proteins in the flipping process, with the trapping data suggesting that the methyltransferases do facilitate flipping.

Klimasauskas et al. have used ^{19}F NMR experiments combined with gel mobility experiments to visualize various conformers during the base-flipping process in the binary (M.*Hha*I−DNA) and ternary complexes (M.*Hha*I−DNA−SAH).[70] Three different conformers were identified for the target base, 5-fluorocytosine, namely, (A) the base stacked in the DNA, (B) intermediate flipped out forms, and (C) the target base binding to the catalytic pocket of the enzyme. In the binary complex, intermediate flipped out forms of the base were observed and addition of the coenzyme to the binary complex was shown to greatly enhance the binding of the target base to the catalytic pocket of enzyme. These results further indicate that the enzyme plays an active role during the base-flipping process.

2-Aminopurine has been shown to be useful as a fluorescent probe for DNA base flipping.[71] The fluorescence of this moiety is quenched when stacked inside the helix but is enhanced dramatically when it is flipped out, due to the loss of stacking interactions with the neighboring bases. Stivers et al., among others, have used a stopped-flow fluorescence kinetic technique by which the fluorescent analogues 2-aminopurine and tryptophan in the enzyme are observed to investigate the base-flipping mechanism.[20,21,72−75] In these studies, 2-aminopurine is positioned next to the flipping base and when the base flips the stacking interaction between the flipping base and 2-aminopurine results in an increase in the fluorescence. The fluorescence due to 2-aminopurine increases by up to 10-fold, while that of tryptophan is decreased by 2-fold enabling observation of real time dynamics of base flipping. The change in the kinetics of base flipping and DNA repair upon mutation has been studied extensively using this technique. Three distinct steps involved in the base-flipping process by uracil DNA glycosylase, namely, (i) formation of the complex, (ii) initiation of flipping, and (iii) binding

of flipped base to the active site, were identified, and the corresponding rate constants were calculated.[76]

2.3.2. Structural Considerations

Structural studies of base flipping have been dominated by X-ray crystallographic analysis of DNA−protein complexes, with the first significant breakthrough being the ternary structure of M.*Hha*I showing the base to indeed be flipped out of the helix into the active site of the enzyme.[11] In addition, X-ray crystallography on DNA alone, NMR, and fluorescence studies have yielded additional insights into the base-flipping phenomena. Highlights from these studies will be presented in the remainder of this section.

On the basis of the ternary structure of M.*Hha*I, a three-step mechanism was proposed, which assumed an active role of the enzyme in flipping: (A) recognition of the target base pair, (B) increase of the interphosphate distance thereby weakening the base pair interactions, and (C) initiation of the base flipping by invasion of the protein and binding of the flipped base in the active site. In the X-ray structure of the M.*Hha*I ternary complex, Gln237 binds to the orphan G base of the DNA through the major groove. On the basis of the crystal structure, it was inferred that Gln237 pushes the target C base from the major groove forcing it to flip via the minor groove, with Gln237 then binding to the orphan G base. Thus, the location of Gln237 would block the target C base from flipping via the major groove. Following flipping of the target C, a conformational rearrangement of the active site loop occurs whereby it closes around the DNA and locks the flipped base in its catalytic pocket. However, the inferred mechanism from this study was limited by knowledge of only the end states of the flipping mechanism; information on the structural relaxation of both the DNA and the protein upon binding was lacking.

X-ray crystal structures of protein−DNA complexes have been very useful in gaining insights into the mechanistic aspects of base flipping. For example, a detailed analysis of the crystal structure of the human uracil-glycosylase (UDG) bound to a DNA containing uracil indicates that the target base is pushed out of the enzyme and then is pulled into the binding pocket, referred to as the push−pull mechanism.[77,78] Mutational analysis reveals that Leu191 in UDG decreases the binding affinity by 60-fold and was found to be responsible for pushing the uracil base from the minor groove.[72,73,79]

A crystallographic study involving a chemically modified DNA duplex was reported by van Aalten et al., which suggests a major groove-opening mechanism.[80] A high-resolution crystal structure of d(CCAGGCCTGG)$_2$ that included an engineered disulfide cross-link exhibits the opening of the central G toward the major groove. It should be stated that this may not be taken as general evidence for a major groove pathway for base flipping since the cross-link might impose constraints on the DNA to force open the base toward the major groove.

A recent study by Horton et al. reported the crystal structure of the M*Hha*I−DNA complex in which the DNA contains an abasic south-constrained target pseudosugar, which shows that the pseudosugar is trapped approximately halfway along the major groove pathway.[81] This forms the first experimental evidence for the base-flipping process to occur via the major groove in the presence of a protein.

Several mutagenesis experiments to understand the role of various amino acids in M.*Hha*I on the binding and specificity have been reported.[66,82−85] A recent study by Merkiene and Klimasauskas demonstrated that the rate-limiting step in methyl transfer is the dissociation of SAH from the ternary complex or opening of the loop of the protein.[86] Binding studies were done with all possible mutants of Gln237, which indicated that the extent of binding of DNA to the enzyme is greatly affected but the specificity in recognizing the cognate sequence remains unaltered.[82] Mutational studies involving Thr250 showed its role in constraining the conformation of the backbone when the target base is rotated out of the helical position. Mutants with bulky side chains including Asp, Asn, and His increased K_m values for both DNA and SAM to the protein.[66] Mutations were carried out for the amino acids in the SAM binding region, and it was found that the K_m for coenzyme is not affected by mutants although the methylase activity varies.[83] On the other hand, mutation of Val121 by Ala affects base flipping and catalysis.[84] Surprisingly, k_{cat} decreases by a factor of 4 or 5 orders of magnitude. More recently, the effects of replacement of residues distal from the active site by alanine on coenzyme binding, methyl transfer, and product release were studied. Most of these mutants did not affect any of the binding or catalytic properties; replacement of Asp73 and Val282 by alanine increased the SAM binding by 25 times and led to a 4-fold increase in the catalytic activity, respectively.[85]

Clearly, the wide variety of experimental studies on base flipping has yielded a picture of a structural perturbation of DNA and RNA that is essential for their biological functions. Information ranges from detailed structural information on the end states of the flipping process, rates and equilibrium constants concerning the transitions between base-paired and base-flipped states, and sampling of those states as well as indications of possible intermediates involved. While this collection of "snapshots" associated with base flipping has greatly increased our understanding of the flipping process, atomic details of the structural transitions and information on the energetic contributions to the structural perturbation are not readily accessible from these experimental methods. Accordingly, a variety of computational studies that allow such phenomena to be observed directly have been performed on systems where base flipping occurs. Details of these approaches will be presented in the following sections.

3. Computational Methods

3.1. General

Theoretical studies based on empirical force fields allow for biological molecules in their aqueous environment to be investigated at an atomic level of detail.[87,88] For the past decade or so, increasing computer power, the availability of high-quality force fields, efficient simulation algorithms, and more rigorous treatment of long-range interactions have resulted in more accurate modeling of biomolecules. Information from these approaches can include structural events occurring during the flipping process, energetics associated with flipping, and how different structural features of the system contribute to the energetics. In the remainder of this section, a number of the issues associated with the application of computational approaches will be addressed followed by a detailed report on the outcomes from a number of computational studies on flipping. The reader is referred to previously published works on the basic background associ-

ated with empirical force field studies of biomolecules, including molecular dynamics (MD) simulations.[87−90]

3.2. Restraints Used for Flipping of the Base

MD simulations of biological systems are typically limited to the nanosecond time regimen. However, as discussed above, base flipping in DNA alone typically occurs on the millisecond or longer time scale, such that it is generally not possible to simply perform an MD simulation in which flipping will occur spontaneously. Accordingly, it is necessary to apply an external potential that forces the flipping process to occur. Such external potentials, or restraints, can be applied in a number of ways, as discussed below. Of the various restraints used to force base flipping out of the DNA duplex, the center of mass (COM) dihedral angle proposed by Banavali and MacKerell[52] and the restraints employed by Lavery and co-workers[91] have been the most successful.

In one of the earliest computational studies on base flipping, Keepers et al. used a restraint on the N1(pyrimidine)−N3(purine) distance to force base pair opening.[92,93] However, the use of a single distance restraint does not allow for the minor vs major groove-flipping pathways to be sampled independently as well as systematically sampling the fully flipped states. Subsequently, Ramstein and Lavery used an internal coordinate restraint to force opening.[94] The target base rotated around an axis perpendicular to the plane of the target base in its WC base-paired state that passed through the center of the sugar attached to the target base. To study the effect of bending on base opening, a second restraint, which forces the terminal helical axis segments to be tangential to a circle corresponding to a chosen radius of curvature, was applied. In combination, these restraints allowed for flipping to be studied via both the minor and the major grooves as a function of bending of the DNA helix.

Chen et al. have proposed that modifying the backbone ζ-torsion angle may be used to induce base opening.[95] They assessed the correlation of the ζ-torsion angle with the extent of base opening by considering all crystal structures of B-DNA duplexes available in the nucleic acid database (NDB).[96] An additional dihedral angle restraint on the glycosyl linkage was employed to ensure that the plane of the opening base stays parallel to that in its original position in the equilibrium structure. Use of a ζ-torsion angle restraint combined with the restraints on the glycosyl linkage ultimately assumes that only these two degrees of freedom are responsible for flipping. If this is not the case, then the use of such restraints will impose unphysical structural changes on the molecule during flipping. Use of this restraint led to an artificial increase of energy in some regions of the potential energy surface generated (discussed below), suggesting limitations in it applicability.

Lavery and co-workers have defined a new restraint, different from the one previously reported from that group, to force flip the target base (Figure 6).[91] Two unit vectors are defined; one along the glycosyl bond of the opening base and the other along the line joining the C1′ atoms of the two bases in the base pair. The opening restraint uses the projection of the angle, θ, formed by these two unit vectors onto the plane perpendicular to the helical axis. The helical axis is defined as the mean of two unit vectors, one connecting the C1′ atoms of the opening base and those of its adjacent bases and the other unit vector connecting the C1′ atoms of the base pair partner and its adjacent bases in the same strand. The use of this projection ensures that increasing the angle actually flips the base out while avoiding any increase in the angle only due to changes in the inclination of the base. In the WC base-paired state, the initial value of this angle is approximately 55°. The difference in this angle and that corresponding to the open base is taken as the reaction coordinate, and both the minor and the major grooves may be accessed. This restraint has been successfully used in a number of base-flipping studies (see below). A similar restraint was also defined to modify the twist in the DNA helix.

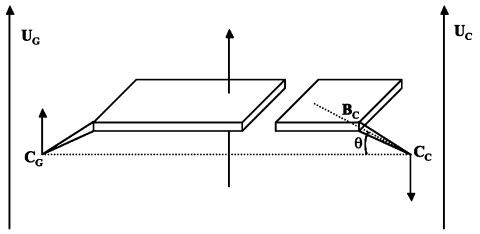

Figure 6. Schematic representation of the base-opening restraint used by Lavery and co-workers. Individual bases of the base pair that undergoes opening are represented by rectangular blocks, and the glycosyl linkage is denoted by the extended triangles attached to them. C_C and C_G are the C1′ atoms of the flipping cytosine and its WC base pair partner guanine, and B_C is the N1 atom of the cytosine base such that B_cC_c is the unit vector along the glycosyl bond. U is the helical axis, which is the mean of U_G and U_C, which connect the C1′ atoms of the adjacent sugar moieties in the corresponding strands. The arrows attached to C_C and C_G show the direction of the individual DNA strands. Base opening is defined as the projection of the angle, θ, formed by B_C, C_C, and C_G, onto the plane perpendicular to U. Adopted from ref 91.

An alternate base-flipping restraint is the COM pseudodihedral angle introduced by Banavali and MacKerell.[52] This restraint allows sampling of both the minor and the major groove base-flipping pathways (Figure 7). The COMs of four sets of atoms were chosen to define the pseudodihedral angle, which is varied to induce base flipping. For example, for flipping of the underlined C base in GCG, the four sets of atoms are the nonhydrogen atoms in (a) the adjacent GC base pair 3′ to the target base, (b) the sugar moiety attached to adjacent G 3′ to the target C base, (c) the sugar attached to the target C base, and (d) the target C base itself. The dihedral angle formed by the centers of mass for the WC base-paired state is approximately 10°, and that for the completely flipped state is ∼190°. Decreasing the angle from 10 to 0 to 190° forces the base to flip via the minor groove, and increasing it from 10 to 190° corresponds to major groove flipping (Figure 7). This restraint, along with the more recent restraint introduced by Lavery and co-workers,[91] may be considered most appropriate for the study of base flipping as they allow for both the major and the minor grooves to be sampled explicitly as well as allowing the internal degrees of freedom of the oligonucleotide to relax freely during the flipping process.

4. Insights into Base Flipping Using Theoretical Approaches

4.1. General

Various experimental studies have been performed to date on base flipping in DNA alone and in the presence of various enzymes, including M.*Hha*I (section 2). While these studies have greatly enhanced our understanding of base flipping, there are inherent limitations in experimental approaches that

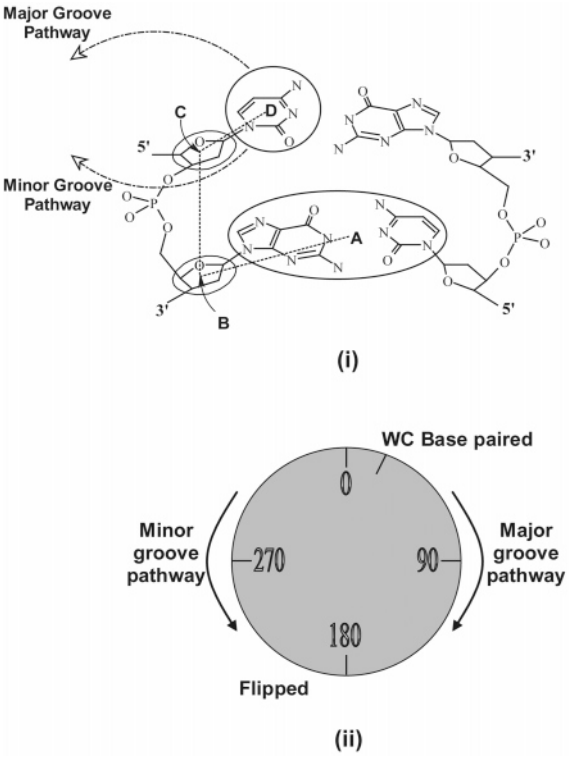

Figure 7. COM pseudodihedral angle base-flipping restraint is defined by the COMs of the atoms incorporated in the four circles A, B, C, and D depicted in (i). A COM dihedral angle of around 10° corresponds to the WC base-paired state for the cytosine base (ii). An angle of approximately 190° corresponds to the fully flipped state. Increasing the COM dihedral angle from 10 to 190° opens the base through the major groove and decreasing it from 10 to 0° and to 190° in the opposite direction is the minor groove pathway.

limit the amount of data that may be obtained. For example, high-energy structures that occur during the flipping process are very short-lived, making them difficult to observe experimentally. In addition, it is difficult to obtain information on the energetic contribution of different regions of the system to the flipping process without significantly perturbing the system (e.g., via mutations). The only viable approach to overcome these limitations and to obtain insights into the atomistic details of the base-flipping process is by computational means. Theoretical calculations at various levels, mainly molecular mechanics (MM) and MD simulations, have been reported. The following subsections provide a detailed report on these computational studies.

4.2. Base Flipping in DNA and RNA Alone

4.2.1. Early Molecular-Modeling Studies

Keepers et al. reported the first MM calculations on the base-opened states in various sequences using an early force field that did not include explicit treatment of the solvent.[92,93] In the study, the purine N3 to pyrimidine N1 distance was forced to a value of 6 Å to emulate the situation in the base-opened state. Initially, energy minimization was done with the distance restraints followed by a minimization with the restraint removed. This resulted in a number of structures with an open conformation for the base, which were approximately 5 kcal/mol higher in energy as compared to the canonical form of the B-DNA. A reaction energy path generated by incremental distance restraints followed by minimization yielded a potential energy barrier of 7−8 kcal/mol as compared to the experimentally observed estimate of 14 kcal/mol. It should be noted that omission of solvent along with the use of minimized structures, such that only potential energies rather than free energies are obtained, severely limited the ability of this approach to perform direct comparisons with experiment. Ramstein and Lavery used an internal coordinate representation of DNA to flip the central thymine in a $(A)_5$ sequence[94] using the FLEX force field[97,98] and the program JUMNA[99] utilizing an implicit solvent model. An additional restraint for bending of the DNA helix also was used. The energy profiles for base opening at various bending angles indicated that these two are strongly coupled, with base opening being facilitated by bending. On the basis of this model, Briki et al. performed a Brownian dynamics simulation to calculate the base pair lifetime, which was found to be in reasonable agreement with the experimental data.[100] The role of helical twist of the DNA in base opening was examined by Bernet et al. by considering the thymine opening in a TA alternating DNA sequence.[91] To account for the solvent environment, two different implicit solvent models were employed, namely, a distance-dependent dielectric function and Poisson−Boltzmann electrostatics. Changes in energy with respect to base opening via either groove assessed using the two different implicit solvent models were qualitatively similar. The helical twist was shown to have a dramatic effect on the base-opening energy profiles especially via the minor groove. The energy change of flipping along the major groove is not significantly altered by untwisting the DNA helix by a degree of 0−40°. Along the minor groove, the energy profile lowers upon untwisting; interestingly, the WC base-paired state was observed to be higher in energy as compared to the base-opened state indicating spontaneous opening. Such a result, as with all of the studies discussed in this article, may be related to the force field being used along with the implicit treatment of solvent in the model.

Chen et al. have generated potential energy profiles for base flipping based on the torsion angle, ζ, as the reaction coordinate.[95,96,101−103] This study explored the energetic pathways from the structural transformation of the canonical form of DNA to the fully flipped state. Parameters were derived from the AMBER force field[104] and supplemented with an explicit term for hydrogen bonding based on a Morse potential, and a distance-dependent dielectric function was used to model the solvent environment. During flipping, the energy via the minor groove increases drastically upon deviating from the WC base-paired state, apparently due to the inherent limitations with the ζ torsion angle restraint discussed above. From these results, it was concluded that base opening through the minor groove is energetically not viable. The energy barrier for base flipping was assessed to be 25.3 kcal/mol as compared to the observed enthalpy of 17−26 kcal/mol for premelting base opening associated with thermal fluctuations.[101] Because of the low energy barrier along the major groove, it was assumed that base flipping happens spontaneously by thermal fluctuations. Moreover, it was argued that flipping of the base might occur before the actual binding of an enzyme. This would enable the DNA to have more recognition sites, and possible steric clashes between the base and the enzyme during base flipping in the presence of protein may be avoided. Correlation of the backbone torsion angles with respect to the base opening showed that, with the exception of the ζ torsion used as the restraint, significant changes did not occur, a result that was

interpreted to validate the use of the ζ torsion to force flipping. They also have compared base opening in RNA and DNA using a similar methodology applying the ζ torsion angle restraint.[103] These calculations indicate that base opening along the major groove in RNA is unfavorable as compared to DNA, which correlates with RNA duplex stability in general. Flipping of adenine, uracil, or cytosine into the minor groove was calculated to be completely forbidden, whereas a guanine base is able to open toward both grooves. Toward the major groove, a maximum opening angle of only 23° could be achieved in RNAs, apparently another limitation of the ζ torsion angle restraint. However, this was suggested to be due to the high thermal stability of RNAs and hence reduced tendency to undergo base opening. The difference in the energetics of the flipping process between DNA and RNA was traced to the helix constraint associated with the A-form.

4.2.2. Free-Energy Calculations of Base Flipping

The computational studies discussed above primarily produced potential energies as a function of the extent of flipping. While these calculations have improved our understanding of the mechanism of base opening, computational approaches that allow for the free energy as a function of the extent of flipping to be obtained may be considered preferable, as they allow for direct comparison with experimental data. In addition, it is important to consider the explicit counterions and solvent environment for a polyanion such as DNA.

As discussed above, to investigate the base-flipping process using MD simulations, it is necessary to introduce an external restraint to force the flipping process to occur. Such restraints may be used in conjunction with MD simulations to calculate the free-energy change along a given reaction coordinate, such as base flipping, an approach termed umbrella sampling. Such a free-energy profile is referred to as a potential of mean force (PMF).[105,106] With respect to base flipping, the umbrella-sampling technique has been employed by MacKerell and co-workers[52,107−109] and Lavery and coworders[53,110−112] using the pseudodihedral (Figure 7) and internal coordinate (Figure 6) restraints, respectively, as the umbrella potentials (i.e., the base-flipping restraints used to define the PMF). These restraints or umbrella potentials are typically applied as a harmonic term of the form

$$\omega_i(x) = k_i(x - x_i)^2 \tag{1}$$

where k_i, x, and x_i are the force constant applied, the actual value of the reaction coordinate, and the restrained value of the reaction coordinate, respectively. Implementation of the umbrella-sampling approach involves performing a series of MD simulations, which include different values of x_i, thereby causing the target base to sample different extents of flipping during the different MD simulations (e.g., windows). The samplings from all of the MD simulations are then combined, yielding a biased probability distribution, $\zeta(x)$, where the bias is associated with the restraints used to enforce the extent of flipping

$$W'(x) = -k_B T \ln \zeta(x) \tag{2}$$

A biased PMF, $W'(x)$, may then be calculated from the probability distribution based on a Boltzmann distribution as shown in eq 2, where k_B is the Boltzmann constant and T is the temperature. The bias in the PMF must then be accounted for as shown in eq 3

$$W_i(x) = W_i'(x) - w_i(x) + F_i \tag{3}$$

where $W_i(x)$ is the unbiased PMF, $w_i(x)$ is the umbrella restraint energy calculated using eq 1, and F_i is a constant for the given window i (i.e., one of the MD simulations used to calculate the PMF). The constant F_i, which effectively connects the windows of the PMF, is typically calculated using the weighted histogram analysis method (WHAM)[113,114] to obtain the unbiased PMF.

The first free-energy profile for the base-flipping process in the context of the GAGAGAGAGAGA sequence using an explicit solvent representation was reported by Guidice et al. using the AMBER program[115,116] with the AMBER Parm99 force field.[117] PMFs were generated for opening of the underlined A and its WC base pair counterpart, T, via both the major and the minor grooves. The calculated free-energy barrier was compared to that from imino proton exchange studies obtained for the opening of the A:T base pair in CAG and GAT sequences. The flipping of the purine base was shown to be easier from the major groove as compared to the minor groove, whereas for the pyrimidine base the change in the free energy is approximately symmetrical along the two pathways. They observed that the free energy increases quadratically near the WC base-paired state, and when the flipping base is moved far enough out of the helix such that base pair and stacking interactions are no longer significant, the energy changes almost linearly. The difference in the behavior of these two types of bases was attributed simply to the size of the purine base, which encounters steric clashes along the minor groove pathway. On the basis of the comparison of the calculated free-energy barrier with the experimental data,[45] it was inferred that imino proton exchange would occur at about ≥50° from the WC state.

M.*Hha*I specifically recognizes the GCGC sequence in DNA and methylates the underlined cytosine in the sequence.[11] Hence, computational studies on this particular sequence were undertaken in our laboratory to shed light into mechanistic aspects of base flipping occurring both alone and in the presence of protein. Initial work involved complete free-energy surfaces for the base-flipping process in the GTCAGCGCATGG sequence.[52] This was performed for both the target C and it's WC partner G and, taking advantage of the COM pseudodihedral restraint described above, included sampling of all of the flipped states via both the minor and the major grooves as well as those in the vicinity of the WC state. Calculation of the PMFs involved MD simulations using the all-atom CHARMM27 nucleic acid force field[118,119] in the presence of an explicit solvent environment using the CHARMM program.[120,121] Figure 8 gives the free-energy profiles obtained for the target C and G flipping. In contrast to the earlier computational studies discussed above, the major and minor groove pathways are comparable in terms of the energetics. The barriers for flipping along the major and minor groove of the DNA for the C base were computed to be 15.3 and 17.6 kcal/mol, respectively, and those for the G base flipping are 21.3 and 18.7 kcal/mol. On the basis of these energy barriers, it was concluded that base opening in DNA could occur via either pathway. Once the base moves out of the low-energy region

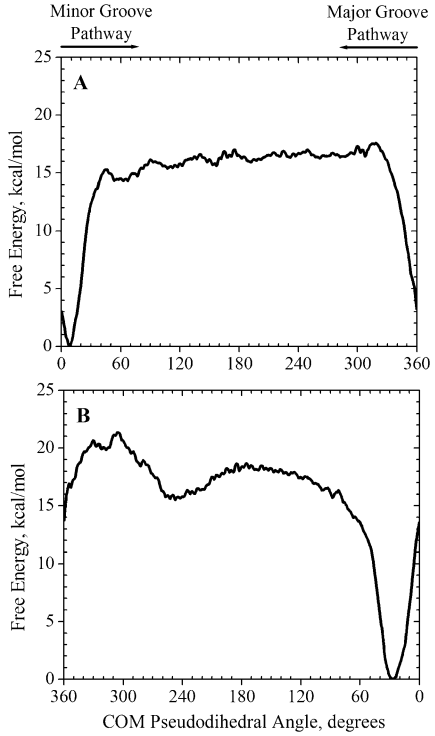

Figure 8. Free-energy profiles calculated for the flipping of target C (A) and G (B) base in the GCGC sequence by Banavali and MacKerell.[52] The WC base-paired state corresponds to a COM dihedral angle of ~10° for the C base and ~30° for the G base. The directions of the major and minor groove pathways are designated in the figures, and the *X*-axis of parts A and B are reversed to allow for minor vs major groove flipping to be in the same directions on the graphs (see original text).[52]

in the vicinity of the WC state, the profile is reasonably flat indicating that the flipped base samples various conformations.

Results from that study were directly compared with the experimental equilibrium between the open and the closed states of flipping base pair from imino proton exchange experiments. Analysis of the solvent accessibility of the G imino proton for both C and G flipping indicated it to be accessible approximately ±30° from the WC base-paired state. Using that range to define the closed state (i.e., assuming that open states correspond to when the imino proton is accessible to solvent) and via conversion of the PMFs from Figure 8 back to probability distributions, the equilibrium between the closed (i.e., WC states) and the open states was calculated. Using two different PMFs based on different amounts of MD sampling, the equilibrium constant was calculated to be 3.3×10^{-8} and 8.8×10^{-8}. These values are in good agreement with the experimental values of 3.3×10^{-7} in a GCGC-containing sequence,[49] validating the computational approach and force field used to calculate the free energies associated with flipping.

Varnai and Lavery reported the free-energy profile of the target C base flipping in GTCAGCGCATGG.[112] Comparison of their results with those of Banavali and MacKerell showed qualitative agreement.[52] However, the overall shapes of the free-energy curves are quite different (Figures 8A and 9). The free energy rises rather sharply along the minor groove as compared to the major groove in the Varnai and Lavery work, which is consistent with the conventional wisdom that base opening along this pathway would be unfavorable due to steric clashes of the exocyclic groups with the WC partner

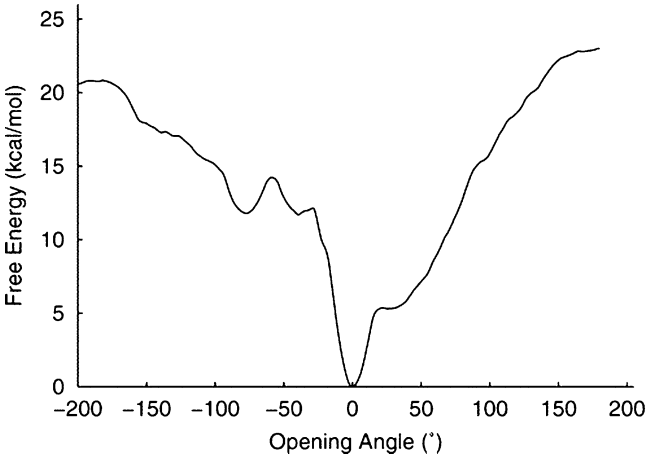

Figure 9. Free-energy changes for target C base flipping as a function of the base-opening angle from the study by Varnai and Lavery.[112] Positive and negative angles denote the major and minor groove pathways, respectively. Reprinted with permission from ref 112. Copyright 2002 American Chemical Society.

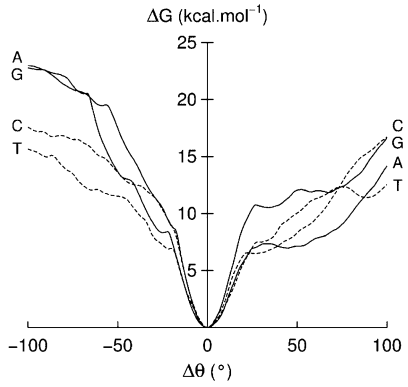

Figure 10. Free-energy profiles corresponding to AT and GC base pair opening in an alternating GA sequence with respect to the relative of the opening angle by Guidice et al.[53] Major and minor groove pathways correspond to the positive and negative values of $\Delta\theta$. Free-energy changes for the opening of purine and pyrimidine bases are represented by solid and broken lines, respectively. Reprinted with permission from ref 53. Copyright 2003 Oxford University Press.

during flipping and the proximity of sugar−phosphate on the minor groove side. The free-energy profile obtained by Banavali and MacKerell is characterized by a flat surface after the free-energy barrier is overcome from either pathway,[52] whereas in the Lavery study, the free energy increases gradually upon going from the WC base-paired state to the fully flipped state (Figure 9). The quantitative differences in the conclusions arrived at by these two studies may be traced to the different force fields used, methodological differences, and type of restraints employed. Ongoing studies in our laboratory are addressing this question.[122]

PMF calculations on flipping by Guidice et al. have compared the conformational pathways and free-energy variations for GC and AT base opening within a B-form DNA d(GAGAGAGAGAGAG)·d(CTCTCTCTCTCTC).[53,111] Similar to their previous studies, both GC and AT base pairs begin to open with a quadratic energy regimen corresponding to elastic deformation, followed by a nearly linear region once the hydrogen bonds of the WC base pairs are broken (Figure 10). For the larger purine bases, G and A, the major groove pathway is clearly favored over the minor groove side, whereas for the pyrimidine bases, both pathways are

energetically comparable. Interesting conformational variations were encountered along the flipping profile. As has been observed previously,[52] the initial movement of the target base from the helix is coupled with the movement of the WC paired partner; this coupled motion is strongest in flipping of T and weakest for C. When the cytosine moves out of the helix, the thymine 3′ to the target base moves along with the flipping base maintaining the stacking interactions. Weakening of the adjacent AT base pair during C flipping was verified by additional calculations on the base opening of thymine by constraining the cytosine in a partially open position. The presence of counterions in the vicinity of opening base alters the free-energy profile significantly; along the major groove, counterions stabilize the open state as demonstrated by a local free-energy minimum.

PMF studies of base flipping have yielded a better understanding of the sequence dependence of the flipping process. On the basis of structural analysis of the flipping process, Banavali and MacKerell suggested a novel mechanism by which the local sequence can impact the base-flipping energy pathway.[52] In addition to the known difference in base stacking and solvent effects, hydrogen-bonding interactions of the target base with the adjacent bases once it is flipped were proposed as possible factors, which might influence the sequence dependence of flipping. Free-energy profiles upon base substitutions in the first three positions of GCGC (i.e., ACGC, GTGC, and GCAC) were calculated in DNA alone by Huang and MacKerell as part of a study on sequence recognition by M.*Hha*I (see below).[109] Substitution of the bases leads to a major change in the calculated PMFs (see Figure 13 below). Near the WC base-paired state, the increase in the free energy due to base flipping is similar for the different sequences. However, moving further along the major groove, the free-energy barriers are significantly lower for ACGC, GTGC, and GCAC as compared to GCGC, while they are higher along the minor groove path. Guidice and Lavery have compared the T base flipping in AAAA and GAGA sequences.[111] Experimental data suggest that the AT base lifetime in A-tracts, which induce DNA bending, is much longer as compared to that observed normally.[47,123] The free-energy profiles calculated for these two sequences are consistent with the experimental observation. It was indeed more difficult to open thymine within the A-tract; the difference in the barriers for these two sequences was calculated to be around 2 kcal/mol toward both of the grooves. The unusual long lifetime of the AT base pair in A-tracts was traced to the additional constraints imposed by a narrow and rigid minor groove. These results have started to yield a better understanding of the impact of sequence on base flipping; further studies are required to better quantify the phenomena.

The number of flipping studies on RNA, including both experimental and theoretical approaches, is minimal. However, the studies that have been performed point toward some significant differences with respect to DNA. NMR imino proton exchange studies indicate that the lifetimes for the d(AT) and AU base pairs are on the order of 5−10 and ≤1 ms, respectively, with the results for the AU base pairs in RNA being an upper bound.[50] Thus, flipping in RNA AU pairs appears to be an order of magnitude or more faster than in DNA AT pairs. In contrast, with GC pairs, the base pair lifetime in RNA is longer than that in DNA. Thymine and uracil base openings in B-DNA and A-RNA in an alternating GA sequence have been studied.[111] Opening of uracil via the major groove is favorable as compared to the minor groove by 1 kcal/mol as is the case of thymine opening in DNA. The free-energy profiles for the two are almost identical; however, major differences in the structural features were observed. In the case of DNA, base opening is coupled with bending, but in RNA, the major groove width increases up to 8 Å and is only 4 Å along the minor groove.

Adenosine deaminases act on RNA and convert adenosine to inosine, which is understood to occur after the adenine base undergoes flipping to assume an extrahelical form. Recently, Nilsson and co-workers have reported free-energy calculations for base flipping in RNA to investigate the selective deamination of the mammalian glutamate receptor B pre-mRNA (gluR-B).[124] GluR-B has two sites containing AC mismatches of which adenosine belonging to only one of these two base pairs undergoes deamination. To investigate the selectivity of deamination, they have performed MD simulations on the RNA and PMF calculations using the methodology initially reported by Banavali and MacKerell.[52] Free-energy profiles for base opening of adenines in the two different parts of the RNA were generated, and the preference of deamination at one site over the other was successfully explained.

Observations consistent with the enhanced opening in RNA AU pairs were obtained in simulation studies of a series of RNA and DNA duplexes.[125] In that study, which included simulations of deoxy sequences containing uracil to control for the base vs 2′-hydroxyl changes, it was observed that local opening events occur into the major groove in RNA, events that may be responsible for the increased opening rate observed via imino proton exchange experiments discussed above. Structural analysis attributed this enhanced opening rate to be due to a "conformational switch" in RNA associated with restrained flexibility about the glycosyl linkage and of the sugar as compared to DNA along with 2′OH(n)-O4′(n + 1) hydrogen bonding. This model was shown to be consistent with other experimental data; however, the calculations also predicted enhanced opening with the GC pairs in RNA, inconsistent with experiment and suggesting a possible force-field bias. It would be of interest to determine if structural events associated with the proposed conformational switch are observed in the PMF-based studies of flipping in RNA.

Osman and co-workers have reported the free-energy surface for base flipping in damaged and undamaged DNA.[126,127] These studies were aimed at understanding the recognition of damaged DNA by repair enzymes such as glycosylases. The ease of base flipping in damaged DNA was compared to that in undamaged DNA. MD simulations using AMBER4.1[116] with an explicit solvent environment were performed on the two conformers, namely, the low-energy canonical form and the completely open form. Two-dimensional PMFs were obtained along the opening angle of the flipped base and bending angle of the DNA helix using the data from the two simulations (i.e., umbrella sampling was not performed). The energy barriers for the base flipping (adenine in A:T and uracil in a GU mismatch) in damaged DNA were calculated to be 9.9 and 9.2 kcal/mol, whereas the corresponding barriers in undamaged DNA (adenine in A:T and cytosine in G:C) are 12.4 and 20.8 kcal/mol (Figure 11). This is expected to increase the rate of base flipping in damaged DNA. On the basis of this, it was concluded that the rate-limiting step is not flipping of the base but the insertion of the base into the catalytic pocket, conformational

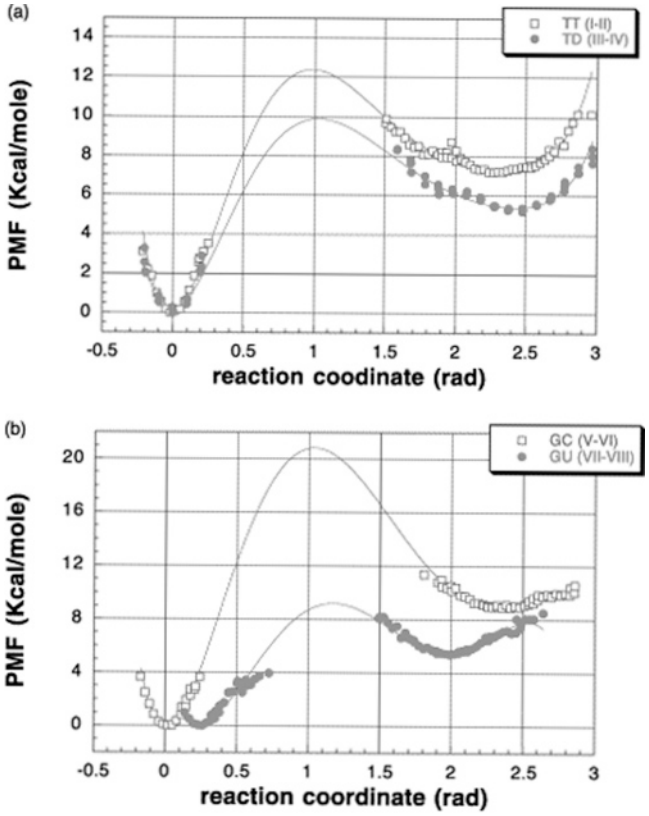

Figure 11. Free-energy profiles by Osman and co-workers[126,127] fitted for the base-flipping process in (a) TT and TD, thymine dimer-containing sequences and (b) GC- and GU-containing sequences. Data obtained from the MD simulations are presented as symbols, and the lines represent polynomial fits to the data. Reprinted with permission from ref 127. Copyright 2002 Elsevier Science Ltd.

changes in the protein, or the cleavage of the glycosyl bond. They also studied the sequence dependence of base pair opening by considering four different damaged DNAs containing G:U wobble base pairs: TGT/AUA, CGC/CUG, and where the central guanine was replaced by 6-methylisoxanthopterin (M) to obtain the other two sequences. The base pairs were observed to flip open spontaneously via the major groove during the MD simulations. On the basis of the population of the open states during the simulations, the order of the difficulty of base flipping was predicted to be TMT < CMC < TGT < CGC. The relative differences in the base opening are explained based on the ability of the G:C base pair to have enhanced stacking interactions with the central G:U base pair as compared to the A:T base pair and the reduced stacking interaction of the 6-methylisoxanthopterin analogue.

A recent study combined experimental and theoretical approaches to investigate the opening mechanism of G:T and G:U mismatches in DNA and RNA, respectively.[54] A two-dimensional PMF was generated for the simultaneous base opening of both G and T. The reaction coordinate used was the linear combination of the opening angles of T and G; the opening of the bases was induced by coupled rotation toward the major/minor groove or by a counterrotation of each of these bases toward a distinct groove. The lowest free energy corresponds to opening of both G and T toward the major groove. At an opening angle of 20 and 10° for T and G, respectively, the free-energy profile corresponds to a shallow minimum, and at angles 40 and 20°, the hydrogen-bonding interactions between the two bases are lost and the imino proton of G is exposed to the solvent and is available for proton exchange. The free energy corresponding to this conformer is calculated to be only 6 kcal/mol as compared to the low-energy canonical form. G and T opening both toward the minor groove or one toward the minor groove and the other toward the major groove are energetically unfavorable and require the opening angles to be higher for the proton exchange to take place. G:U base pair opening yielded similar results, and the free energies calculated are within the error of the simulation protocol. Comparison of the free-energy cost for G:C opening suggests that the opening of the bases toward the major groove remains the lowest free-energy pathway, but the barrier is higher than G:T base opening by about 2.5 kcal/mol. In terms of structure−energy correlation, T(H3)−G(O6) hydrogen bond breaking corresponds to a free energy of 2.5 kcal/mol and G(H1)−T(O2) bond breaking occurs at a free-energy change of 7 kcal/mol. This is consistent with the experimental finding that the T(H3)−G(O6) hydrogen bond is short-lived as compared to G(H1)−T(O2).

4.3. Base Flipping in the Presence of Proteins

4.3.1. Protein Facilitated Base Flipping

As discussed in the Experimental Section, the role of proteins in facilitating base flipping has been a subject of discussion, with conclusions based on experimental rate constants that vary significantly dependent on the data used. To address the question of a passive versus an active role directly, Huang et al. performed PMF calculations on the flipping of DNA in the presence of M.*Hha*I.[107−109] These calculations have indeed shown that the protein does play an active role in base flipping, as well as identified the flipping pathway in the presence of M.*Hha*I and elucidated the mechanism by which the protein recognizes specific DNA sequences.

To determine if M.*Hha*I plays an active role in base flipping, PMFs for base flipping in the DNA dodecamer (GTCAGCGCATGG) in four different environments were generated as follows: (a) DNA alone; (b) DNA in a binary complex with M.*Hha*I in its open conformation, where the active site loop of the enzyme is in the extended conformation; (c) DNA in a binary complex with M.*Hha*I, where the active site loop is closed around the DNA; and (d) DNA in a ternary complex with M.*Hha*I in its closed conformation and the coenzyme, S-adenosylhomocysteine. All of the simulations were performed in the presence of explicit solvent and ions, and the PMF calculations were performed using a similar methodology to that reported by Banavali and MacKerell but used stochastic boundary conditions instead of periodic boundary conditions.[52]

Figure 12 depicts the change in the free energies as a function of the COM pseudodihedral angle, where the DNA is in the four distinct environments. Expectedly, there exist distinct barriers for base flipping via both the major and the minor grooves while the minimum energy structure corresponds to the WC base-paired state. Flipping of the target base is not facilitated by the protein in the binary complex with the open form of the active site loop (dashed line). Interestingly, the free energies associated with the fully flipped states are higher as compared to those for the DNA in aqueous solution, while the free energies associated with the conformers around the WC base-paired state are very similar. However, when M.*Hha*I assumes the closed form

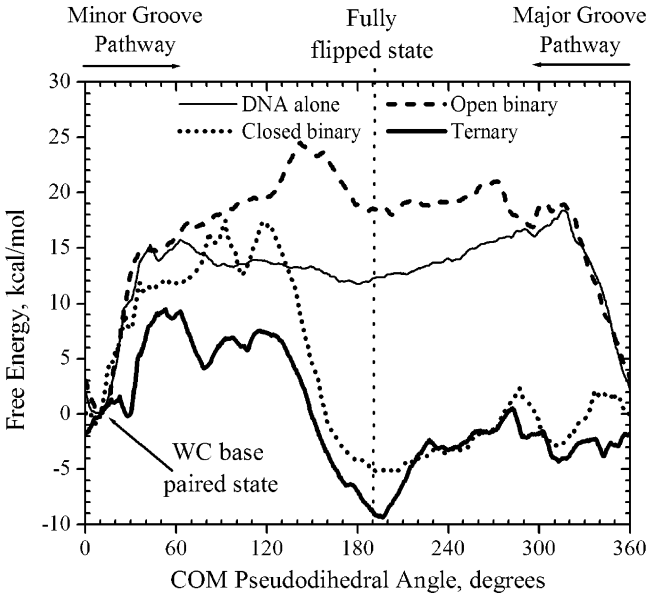

Figure 12. Free-energy profiles for the base flipping of GCGC in four different environments from Huang et al.:[107] DNA alone (thin line), binary complex of DNA and the open form of M.*Hha*I (dashed line), closed binary complex of DNA and M.*Hha*I in its closed form (dotted line), and ternary complex of DNA, SAH, and M.*Hha*I in its closed form (thick line). The free energies were depicted in reference to that obtained for the WC base-paired state (COM pseudodihedral angle of 10°).[107]

of the active site loop in either the presence or the absence of cofactor, there is a huge impact on the free-energy profiles (dotted and thick lines in Figure 12). The barriers for flipping from the major groove were calculated to be 2.5 and 0.4 kcal/mol for the closed binary and ternary complexes, respectively. Thus, PMF calculations conclusively showed that M.*Hha*I does actively facilitate base flipping and that flipping occurs via the major groove in contrast to assumptions of a minor groove path made previously (see above).

Structural analysis of MD simulations used for calculation of the PMF indicated the atomic details by which M.*Hha*I facilitates flipping. First, the major groove pathway involves the flipping base moving through the protein matrix. While this would intuitively be considered to be energetically unfavorable, the protein can readily relax around the flipping base and, importantly, shields the base from the aqueous environment; exposure of the base to the aqueous environment is energetically unfavorable as evidenced by PMF calculations on flipping in DNA alone. Initiation of flipping was indicated to be associated with destabilization of the ground-state WC base-paired structure, where the protein disrupts the normal WC hydrogen-bonding and base-stacking interactions that occur in the normal double helical conformation. In addition, as the base conformation moves toward the fully flipped conformation, protein−DNA interactions replace the lost WC and stacking interactions. Finally, interactions of the protein with the phosphodiester backbone are suggested to be important for flipping,[108] which may contribute to ground-state destabilization. Thus, M.*Hha*I facilitates base flipping via a combination of ground-state destabilization, exclusion of the flipping base from the aqueous environment, and replacement of lost WC hydrogen-bonding and base-stacking interactions with protein−DNA interactions.

A recent combined experimental and computational study by Horton et al. supports a major groove base-flipping pathway in the presence of M.*Hha*I.[81] In that study, an abasic sugar analogue that restrains the pucker to the south conformation was substituted into the DNA at the normal target C site and the crystal structure was determined. Such an oligonucleotide has previously been shown to bind preferentially over an abasic site to M.*Hha*I.[128] In the crystal structure, it was observed that the constrained sugar assumed a conformation that was approximately halfway along the major groove-flipping path, thereby lending experimental support to the major groove pathway proposed by the PMF calculations. In addition, on the basis of the X-ray structure, binding affinity data, and MD simulations, it was suggested that the constrained sugar mimicked the transition state for the flipping of the sugar out of the DNA duplex.

4.3.2. Energetic Recognition of Specific Sequence of DNA by M.HhaI

Methyltransferases have to recognize the correct sequence for methylation, otherwise called the cognate sequence. For example, M.*Hha*I recognizes the GCGC sequence in DNA and selectively methylates the cytosine at the second position. However, on the basis of the crystal structure of the M.*Hha*I−protein complexes, the nature of the specific recognition was not evident, leading to the suggestion that recognition and base flipping are intimately linked.[11] In terms of energetics, this suggests a model where the transition state barrier for base flipping in the cognate sequence is lower as compared to that in noncognate sequences, such that flipping can only occur with the cognate sequence. To address this directly, free-energy profiles (Figure 13) were generated for base flipping in DNA alone, the closed binary, and the ternary complexes by substituting the GCGC sequence at the first three positions, yielding ACGC, GTGC, and GCAC, on which PMF calculations were performed.[109] For the DNA-alone systems, the barriers for the flipping via the major groove were less for the noncognate sequences as compared to the cognate sequence, indicating that the sequence specificity was not associated with flipping in DNA alone. However, in the closed binary and ternary complexes, distinct differences in the PMFs were observed. In both complexes, two distinct free-energy barriers were observed along the major groove for the noncognate sequences, one in the vicinity WC-paired state ($x = 345°$) and the other at $x = 285°$, in contrast to the cognate sequence, where the barriers were close to zero. Thus, the calculations support the hypothesis of an energetic mechanism for sequence recognition by M.*Hha*I.

A detailed analysis of the trajectories from the PMF calculations yielded information on the recognition mechanism. The change in the free energies in the vicinity of the WC state indicates that recognition is associated with two barriers between the WC and the flipped state. The first barrier is directly adjacent to the WC state and is relatively small. Its location is such that in two of the noncognate sequences destabilization of the WC interactions occurred as seen in the cognate sequence, although a barrier to flipping is soon encountered. Beyond that barrier, a second larger barrier is present in all of the noncognate sequences. Thus, in noncognate sequences, the enzyme is predicted to destabilize the WC base-paired state in some cases; however, two barriers to flipping, referred to as selectivity filters, one adjacent to the WC state, and one approximately halfway along the flipping pathway are then encountered, thereby disallowing flipping. Energetic and structural analysis of the

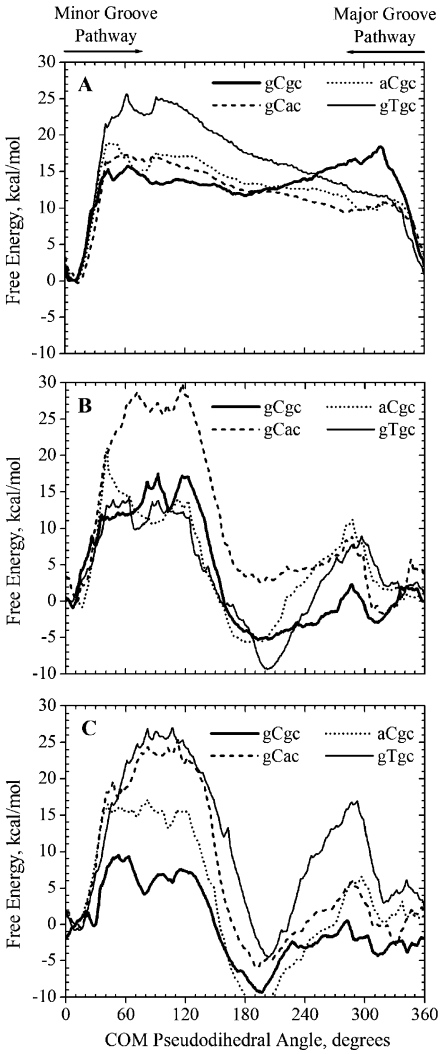

Figure 13. Free-energy profiles for the base flipping of the second base in the four sequences gCgc (thick line), aCgc (dotted line), gCac (dashed line), and gTgc (thin line) for DNA alone (A), the closed binary (B), and the ternary (C) complexes. The free energies depicted are relative to those obtained for the WC base-paired state.[109]

different PMFs in the vicinity of the barriers to flipping showed a number of amino acids to contribute to the barriers to flipping, yielding a general model where the selectivity filters are associated with a "web of interactions" between the protein and the DNA. Such a web of interactions is important for the specific recognition of the entire GCGC sequence by the protein. Subsequently, a combined crystallography/theoretical study on 8-oxoguanine DNA glycosylase I indicated that enhanced DNA−protein interactions are responsible for the specificity of the glycosylase for the damaged base,[62] consistent with our earlier observations on the mechanism of specificity.

4.3.3. MD Simulations on M.HhaI

MD simulations have been performed on the ternary complex of M.*Hha*I by Lau and Bruice to study the dynamics of the active site.[129] These simulations do not concentrate on base flipping; however, interesting observations were made in the region of the active site. The flipped cytosine is reasonably rigid within the active site; however, fluctuations in the torsion angle χ bring about changes in the distance between the SG of Cys81 and C6 of the cytosine and the methyl carbon of SAM and C5 of cytosine. These structural fluctuations are crucial for formation of the covalent intermediate during methylation and are correlated to transition state structures obtained using ab initio calculations. On the basis of the simulations, they have suggested that the source for the protonation of N3 of the target base, a step involved in methyl transfer, may be either from the protonated amine group of SAM or from Arg163 via a water bridge. This suggestion is in contrast to the currently accepted catalytic mechanism discussed above (Figure 2). The conserved residues, Asn304 and Gln82, in methyltransferases seem to stabilize a water network in the active site of the enzyme, which can easily accept the proton.

Wang et al. have reported MD simulations accompanied by experimental studies where the flipped cytosine in the DNA is substituted with an abasis furanose sugar or conformationally constrained abasic pseudosugar analogues that mimic the south and north conformations.[128] On the basis of experimental binding affinities and gel shift data, it was indicated that the abasic south-constrained sugar forms a closed complex with the enzyme, leading to the tighter binding. This was confirmed by MD simulations of the three DNA−M.*Hha*I complexes. In addition, ab initio quantum mechanical calculations suggested that the difference in binding of the north- vs south-constrained sugars was associated with differences in the intrinsic flexibility of the phosphodiester backbone due to the constrained sugars.

4.4. Validation of the Computational Studies

From the above discussion, it is clear that computational studies have yielded novel atomic-detail insights into base flipping and have enhanced our understanding of the base-flipping structural and energetic pathways. However, it is also evident that the computational methods can yield somewhat contrasting results. These differences may be affected largely by (A) the force field used, (B) the restraint methodology used for flipping of the base, and (C) simulation protocol. For example, Banavali and MacKerell[52] and Varnai and Lavery[112] reported C base flipping in the same sequence. While some similarities occurred in the vicinity of the WC states, the overall shapes of the free-energy profiles were distinctly different owing to the above reasons. Nucleic acid force fields are parametrized based on the experimental information on the canonical structures of nucleic acids and high level ab initio calculations on model compounds in primarily low-energy conformations. In the base-flipping process, various high-energy conformations of the oligonucleotide are encountered upon deviating from the WC base-paired state. Thus, to properly model the structural transformations involving high-energy states, the force field used has to accurately treat both the minimum energy and the higher energy conformations. In addition, because the base moves from the interior of the DNA double helix to a solvent-exposed orientation, it is essential that the force field properly reproduces the relative strengths of WC hydrogen bonding, base stacking, and nucleic acid−water interactions to be able to correctly model the flipping process.

Validation of the ability of the simulation model requires that it reproduces available experimental data. As discussed above, CHARMM27[118,119] has been shown to yield near quantitative agreement for the equilibrium between the open and the closed states with respect to flipping of GC pairs.[52] Recently, we have extended this approach to compare the performances of the three popularly used nucleic force fields,

CHARMM,[118,119] AMBER,[104] and BMS.[130] Comparison with the experimental data indicates that CHARMM gives excellent agreement with the imino proton exchange data closely followed by AMBER, whereas BMS deviates considerably.[122] Notably, the individual contributions (e.g., base stacking, WC hydrogen bonding) to the PMFs were significantly different for the three models, emphasizing the impact of the force field on the atomic detail picture obtained from PMF calculations.

Other experimental data can be used to validate the applied simulation protocols and force fields. For example, the base-flipping studies in the presence of M.*Hha*I yielded structures of the fully flipped conformation of the DNA−M.*Hha*I−SAH ternary complex in excellent agreement with the crystal structure. In addition, the flipping PMF results are consistent with data from enzyme kinetic studies. The fully flipped state is more stable relative to the destabilized WC base-paired state by about 5.1 kcal/mol in the closed binary complex and by 9.4 kcal/mol in the ternary complex. This difference in the stabilities upon going from the binary to the ternary complex is in excellent agreement with the experimental work, which indicated that SAH binding enhances the binding of DNA to M.*Hha*I by 4 kcal/mol.[65]

In general terms of simulation conditions, explicit inclusion of water and counterions and proper treatment of long-range interactions are essential. Inadequate sampling length can lead to inaccuracy in the obtained results, requiring that careful convergence of the computed results be verified. The dependence of the nature of the restraint used on the energy/free-energy profiles seems to be important. Care must be taken to ensure that this restraint does not lead to the sampling of high-energy regions so that the energy profiles are close to the experimental regimen. Once these precautions have been taken, it is still necessary to compare the calculated results with available experimental data to further validate the applied methods.

5. Summary and Future Prospects

Enhancements in computational approaches, including more accurate MD simulations, have enabled the study of a wider range of biological systems, including DNA, RNA, and DNA−protein complexes involved in base flipping. While experimental studies formed the basis for establishing base flipping as a biologically relevant event, computational approaches have been the ultimate choice in unraveling the atomic level details of the base-flipping mechanism in both the presence and the absence of proteins. The role of the enzyme in base flipping was debated to be active or passive based on various experimental studies. The active role of protein in facilitating the base-flipping process and mechanistic details of these events have now been elucidated by computational means. Moreover, the specificity of M.*Hha*I in recognizing the correct sequence for methylation was explained by changes in the free-energy barriers to flipping based on PMF calculations. These studies show that computational approaches are instrumental not only in explaining experimental observations but also in gaining insights into problems, which cannot be readily accessed by experiments. Nonetheless, it has to be noted that theoretical approaches have inherent limitations in terms of the methodology adopted, the various approximations used, and the force field applied. While improvements in force fields, including the addition of electronic polarizability[131,132] and increases in computer power, will help to overcome these limitations, validation of the computed results by comparison with experimental data is and will remain crucial in arriving at meaningful conclusions. While computational studies on base flipping may still be considered in their infancy, with only three reports on base flipping in the presence of proteins in the literature, it is anticipated that additional computational studies, including sequence effects and targeting different base-flipping enzymes, will increase our understanding of the base-flipping phenomenon.

6. Abbreviations

COM	center of mass
DNA	deoxyribonucleic acid
MD	molecular dynamics
M.*Hha*I	(cytosine-5)-methyltransferase from *Haemophilus haemolyticus*
MM	molecular mechanics
NDB	nucleic acid database
NMR	nuclear magnetic resonance
PMF	potential of mean force
RNA	ribonucleic acid
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
WC	Watson−Crick

7. Acknowledgments

Financial support is acknowledged from the NIH (GM 51501), and we thank DOD ACS Major Shared Resource Computing and the Pittsburgh Supercomputing Center for their generous CPU allocations. Special thanks are extended to Drs. Victor Marquez, Judith Christman, Irina Russu, Xiaodong Cheng, and Richard Roberts for many helpful discussions.

8. References

(1) Scharer, O. D. *Angew. Chem.-Int. Ed.* **2003**, *42*, 2946−2974.
(2) Lindahl, T.; Wood, R. D. *Science* **1999**, *286*, 1897−1905.
(3) Bestor, T. H.; Verdine, G. L. *Curr. Opin. Cell Biol.* **1994**, *6*, 380−389.
(4) Nelson, H. C. M.; Bestor, T. H. *Chem. Biol.* **1996**, *3*, 419−423.
(5) Garvie, C. W.; Wolberger, C. *Mol. Cell* **2001**, *8*, 937−946.
(6) Roberts, R. J. *Cell* **1995**, *82*, 9−12.
(7) Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181−198.
(8) Kumar, S.; Cheng, X.; Klimasauskas, S.; Mi, S.; Posfai, J.; Roberts, R. J.; Wilson, G. G. *Nucleic Acids Res.* **1994**, *22*, 1−10.
(9) Cheng, X.; Roberts, R. J. *Nucleic Acids Res.* **2001**, *29*, 3784−3795.
(10) Cheng, X.; Blumenthal, R. M. *Structure* **1996**, *4*, 639−645.
(11) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.
(12) Svedruzic, Z. M.; Reich, N. O. *Biochemistry* **2004**, *43*, 11460−11473.
(13) Vilkaitis, G.; Merkiene, E.; Serva, S.; Weinhold, E.; Klimasauskas, S. *J. Biol. Chem.* **2001**, *276*, 20924−20934.
(14) Schubert, H. L.; Blumenthal, R. M.; Cheng, X. D. *Trends Biochem. Sci.* **2003**, *28*, 329−335.
(15) O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X. *J. Mol. Biol.* **1996**, *261*, 634−645.
(16) O'Gara, M.; Horton, J. R.; Roberts, R. J.; Cheng, X. *Nat. Struct. Biol.* **1998**, *5*, 872−877.
(17) Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995**, *82*, 143−153.
(18) Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. *Nat. Struct. Biol.* **2001**, *8*, 101−103.
(19) Blumenthal, R. M.; Cheng, X. D. *Nat. Struct. Biol.* **2001**, *8*, 101−103.
(20) Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37−65.
(21) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729−2759.
(22) Vassylyev, D. G.; Kashiwagi, T.; Mikami, Y.; Ariyoshi, M.; Iwai, S.; Ohtsuka, E.; Morikawa, K. *Cell* **1995**, *83*, 773−782.
(23) Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. *EMBO J.* **1998**, *17*, 5214−5226.
(24) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. *Nature* **1996**, *384*, 87−92.

(25) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5083−5088.
(26) Barrett, T. E.; Savva, R.; Panayotou, G.; Barlow, T.; Brown, T.; Jiricny, J.; Pearl, L. H. *Cell* **1998**, *92*, 117−129.
(27) Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, E. *Cell* **1998**, *95*, 249−258.
(28) Hollis, T.; Ichikawa, Y.; Ellenberger, T. *EMBO J.* **2000**, *19*, 758−766.
(29) Thayer, M. M.; Ahern, H.; Xing, D. X.; Cunningham, R. P.; Tainer, J. A. *EMBO J.* **1995**, *14*, 4108−4120.
(30) Hosfield, D. J.; Guan, Y.; Haas, B. J.; Cunningham, R. P.; Tainer, J. A. *Cell* **1999**, *98*, 397−408.
(31) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. *Nature* **2000**, *403*, 451−456.
(32) Hornby, D. P.; Ford, G. C. *Curr. Opin. Biotechnol.* **1998**, *9*, 354−358.
(33) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33−38.
(34) Ahmad, I.; Rao, D. N. *Crit. Rev. Biochem. Mol. Biol.* **1996**, *31*, 361−80.
(35) Lloyd, R. S.; Cheng, X. D. *Biopolymers* **1997**, *44*, 139−151.
(36) Gromova, E. S.; Khoroshaev, A. V. *Mol. Biol. (Moskow)* **2003**, *37*, 300−314.
(37) Mol, C. D.; Parikh, S. S.; Putnam, C. D.; Lo, T. P.; Tainer, J. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 101−128.
(38) David, S. S.; Wiliams, S. D. *Chem. Rev.* **1998**, *98*, 1221−1261.
(39) Sankpal, U. T.; Rao, D. N. *Crit. Rev. Biochem. Mol. Biol.* **2002**, *37*, 167−197.
(40) Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383−413.
(41) Guéron, M.; Kochoyan, M.; Leroy, J.-L. *Nature* **1987**, *328*, 89−92.
(42) Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Gueron, M. *J. Mol. Biol.* **1988**, *200*, 223−238.
(43) Moe, J. G.; Folta-Stogniew, E.; Russu, I. M. *Nucleic Acids Res.* **1995**, *23*, 1984−1989.
(44) Chen, C. J.; Russu, I. M. *Biophys. J.* **2004**, *87*, 2545−2551.
(45) Foltastogniew, E.; Russu, I. M. *Biochemistry* **1994**, *33*, 11016−11024.
(46) Kochoyan, M.; Leroy, J. L.; Gueron, M. *Biochemistry* **1990**, *29*, 4799−4805.
(47) Snoussi, K.; Leroy, J. L. *Biochemistry* **2001**, *40*, 8898−8904.
(48) Leroy, J.-L.; Bolo, N.; Figueroa, N.; Plateau, P.; J., G. M. *Biomol. Struct. Dyn.* **1985**, *2*, 915−939.
(49) Dornberger, U.; Leijon, M.; Fritzsche, H. *J. Biol. Chem.* **1999**, *274*, 6957−6962.
(50) Snoussi, K.; Leroy, J. L. *Biochemistry* **2002**, *41*, 12467−12474.
(51) Leijon, M.; Graslund, A. *Nucleic Acids Res.* **1992**, *20*, 5339−5343.
(52) Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *319*, 141−160.
(53) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434−1443.
(54) Varnai, P.; Canalia, M.; Leroy, J. L. *J. Am. Chem. Soc.* **2004**, *126*, 14659−14667.
(55) Hoffman, J. L.; Bock, R. M. *Biochemistry* **1970**, *9*, 3542−3550.
(56) Darrington, R. T.; Xiang, T. X.; Anderson, B. D. *Int. J. Pharm.* **1990**, *59*, 35−44.
(57) Komiyama, M.; Takeshige, Y. *J. Org. Chem.* **1989**, *54*, 4936−4939.
(58) Spies, M. A.; Schowen, R. L. *J. Am. Chem. Soc.* **2002**, *124*, 14049−14053.
(59) David, A.; Bleimling, N.; Beuck, C.; Lehn, J. M.; Weinhold, E.; Teulade-Fichou, M. P. *Chembiochem* **2003**, *4*, 1326−1331.
(60) Berthet, N.; Michon, J.; Lhomme, J.; Teulade-Fichou, M. P.; Vigneron, J. P.; Lehn, J. M. *Chem. Eur. J.* **1999**, *5*, 3625−3630.
(61) Jourdan, M.; Garcia, J.; Lhomme, J.; Teulade-Fichou, M. P.; Vigneron, J. P.; Lehn, J. M. *Biochemistry* **1999**, *38*, 14205−14213.
(62) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. *Nature* **2005**, *434*, 612−618.
(63) Wu, J. C.; Santi, D. V. *J. Biol. Chem.* **1987**, *262*, 4778−4786.
(64) Cheng, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J. *Cell* **1993**, *74*, 299−307.
(65) Lindstrom, W. M., Jr.; Flynn, J.; Reich, N. O. *J. Biol. Chem.* **2000**, *275*, 4912−4919.
(66) Vilkaitis, G.; Dong, A.; Weinhold, E.; Cheng, X.; Klimasauskas, S. *J. Biol. Chem.* **2000**, *275*, 38722−38730.
(67) Gromova, E. S.; Oretskaya, T. S.; Eritja, R.; Guschlbauer, W. *Biochem. Mol. Biol. Int.* **1995**, *36*, 247−255.
(68) Reich, N. O.; Mashhoon, N. *Biochemistry* **1991**, *30*, 2933−2939.
(69) Allan, B. W.; Reich, N. O.; Beechem, J. M. *Biochemistry* **1999**, *38*, 5308−5314.
(70) Klimasauskas, S.; Szyperski, T.; Serva, S.; Wuthrich, K. *EMBO J.* **1998**, *17*, 317−324.
(71) Holz, B.; Klimasauskas, S.; Serva, S.; Weinhold, E. *Nucleic Acids Res.* **1998**, *26*, 1076−1083.
(72) Jiang, Y. L.; Stivers, J. T.; Song, F. H. *Biochemistry* **2002**, *41*, 11248−11254.
(73) Jiang, Y. L.; Stivers, J. T. *Biochemistry* **2002**, *41*, 11236−11247.
(74) Allan, B. W.; Beechem, J. M.; Lindstrom, W. M.; Reich, N. O. *J. Biol. Chem.* **1998**, *273*, 2368−2373.
(75) Malygin, E. G.; Evdokimov, A. A.; Zinoviev, V. V.; Ovechkina, L. G.; Lindstrom, W. M.; Reich, N. O.; Schlagman, S. L.; Hattman, S. *Nucleic Acids Res.* **2001**, *29*, 2361−2369.
(76) Stivers, J. T.; Pankiewicz, K. W.; Watanabe, K. A. *Biochemistry* **1999**, *38*, 952−963.
(77) Kunkel, T. A.; Wilson, S. H. *Nature* **1996**, *384*, 25−26.
(78) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. *Nature* **1996**, *384*, 87−92.
(79) Jiang, Y. L.; Kwon, K.; Stivers, J. T. *J. Biol. Chem.* **2001**, *276*, 42347−42354.
(80) van Aalten, D. M. F.; Erlanson, D. A.; Verdine, G. L.; Joshua-Tor, L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11809−11814.
(81) Horton, J. R.; Ratner, G.; Banavali, N. K.; Huang, N.; Choi, Y.; Maier, M. A.; Marquez, V. E.; MacKerell, A. D., Jr.; Cheng, X. *Nucleic Acids Res.* **2004**, *32*, 3877−3886.
(82) Mi, S.; Alonso, D.; Roberts, R. J. *Nucleic Acids Res.* **1995**, *23*, 620−627.
(83) Sankpal, U. T.; Rao, D. N. *Nucleic Acids Res.* **2002**, *30*, 2628−2638.
(84) Estabrook, R. A.; Lipson, R.; Hopkins, B.; Reich, N. *J. Biol. Chem.* **2004**, *279*, 31419−31428.
(85) Sharma, V.; Youngblood, B.; Reich, N. *J. Biomol. Struct. Dyn.* **2005**, *22*, 533−543.
(86) Merkiene, E.; Klimasauskas, S. *Nucleic Acids Res.* **2005**, *33*, 307−315.
(87) *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel-Dekker: New York, 2001.
(88) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(89) MacKerell, A. D., Jr. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F. I., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 3.
(90) Tuckerman, M. E.; Martyna, G. J. *J. Phys. Chem. B* **2000**, *104*, 159−178.
(91) Bernet, J.; Zakrzewska, K.; Lavery, R. *J. Mol. Struct. (THEOCHEM)* **1997**, *398*, 473−482.
(92) Keepers, J.; Kollman, P. A.; James, T. L. *Biopolymers* **1984**, *23*, 2499−2511.
(93) Keepers, J. W.; Kollman, P. A.; Weiner, P. K.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5537−5541.
(94) Ramstein, J.; Lavery, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7231−7235.
(95) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *J. Biomol. Struct. Dyn.* **1998**, *15*, 765.
(96) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *Chem. Phys. Lett.* **1998**, *287*, 570−574.
(97) Lavery, R.; Sklenar, H.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.* **1986**, *3*, 989−1014.
(98) Lavery, R.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.* **1986**, *6*, 1155−1170.
(99) Lavery, R. In *Unusual DNA Structures*; Wells, R. D., Harvey, S. C., Eds.; Springer: New York, 1988.
(100) Leroy, J. L.; Broseta, D.; Gueron, M. *J. Mol. Biol.* **1985**, *124*, 165−178.
(101) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *Phys. Rev. E* **2000**, *62*, 1133−1137.
(102) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *Phys. Rev. E* **1998**, *58*, 909−913.
(103) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *Phys. Rev. E* **2000**, *61*, 5640−5645.
(104) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−861.
(105) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(106) Leach, A. R. *Molecular Modelling: Principles and Applications*; Longman: Harlow, 1996.
(107) Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68−73.
(108) Huang, N.; MacKerell, A. D., Jr. *Philos. Trans. R. Soc. London, Ser. A* **2004**, *362*, 1439−1460.
(109) Huang, N.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2005**, *345*, 265−274.
(110) Giudice, E.; Várnai, P.; Lavery, R. *ChemPhysChem* **2001**, *11*, 673−677.
(111) Giudice, E.; Lavery, R. *J. Am. Chem. Soc.* **2003**, *125*, 4998−4999.
(112) Varnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272−7273.
(113) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(114) Crouzy, S.; Baudry, J.; Smith, J. C.; Roux, B. *J. Comput. Chem.* **1999**, *20*, 1644−1658.
(115) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287−303.
(116) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(117) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(118) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.
(119) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(120) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(121) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 1.
(122) Priyakumar, U. D.; MacKerell, A. D. *J. Chem. Theor. Comput.*, published online Nov 16, http://dx.doi.org/10.1021/ct0501957.
(123) Dickerson, R. E.; Goodsell, D. S.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3579−3583.
(124) Hart, K.; Nystrom, B.; Ohman, M.; Nilsson, L. *RNA Publ. RNA Soc.* **2005**, *11*, 609−618.
(125) Pan, Y.; MacKerell, A. D., Jr. *Nucleic Acids Res.* **2003**, *31*, 7131−7140.
(126) Seibert, E.; Ross, J. B. A.; Osman, R. *J. Mol. Biol.* **2003**, *330*, 687−703.
(127) Fuxreiter, M.; Luo, M.; Jedlovszky, P.; Simon, I.; Osman, R. *J. Mol. Biol.* **2002**, *323*, 823−834.
(128) Wang, P.; Nicklaus, M. C.; Marquez, V. E.; Brank, A. S.; Christman, J.; Banavali, N. K.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2000**, *122*, 12422−12434.
(129) Lau, E. Y.; Bruice, T. C. *J. Mol. Biol.* **1999**, *293*, 9−18.
(130) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487−509.
(131) Anisimov, V. M.; Vorobyov, I. V.; Lamoureux, G.; Noskov, S.; Roux, B.; MacKerell, A. D., Jr. *Biophys. J.* **2004**, *86*, 415a.
(132) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. *J. Chem. Theor. Comput.* **2005**, *1*, 153−168.

CR040475Z