Published in final edited form as:

J Chem Theory Comput. 2009 November 10; 5(11): 3105–3113. doi:10.1021/ct9001575.

An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies

Kun Song[†], Arthur J. Campbell[†], Christina Bergonzo[†], Carlos de los Santos[#], Arthur P. Grollman[#], and Carlos Simmerling^{†,§}

[†]Department of Chemistry, Stony Brook University, Stony Brook, NY 11794-3400

[#]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-3400

[§]Center for Structural Biology, Stony Brook University, Stony Brook, NY 11794-3400

Abstract

Base flipping is a common strategy utilized by many enzymes to gain access to the functional groups of nucleic acid bases in duplex DNA which are otherwise protected by the DNA backbone and hydrogen bonding with their partner bases. Several X-ray crystallography studies have revealed flipped conformations of nucleotides bound to enzymes. However, little is known about the base flipping process itself, even less about the role of the enzymes. Computational studies have used umbrella sampling to elicit the free energy profile of the base flipping process using a pseudo-dihedral angle to represent the reaction coordinate. In this study we have used an unrestrained trajectory in which a flipped base spontaneously reinserted into the helix in order to evaluate and improve the previously defined pseudo-dihedral angle. Our modified pseudo-dihedral angles use a new atom selection to improve the numerical stability of the restraints and also provide better correlation to the extent of flipping observed in simulations. Furthermore, based on the comparison of PMFs generated using different reaction coordinates, we observed that the shape of a flipping PMF profile is strongly dependent on the definition of the reaction coordinate, even for the same data set.

1. Introduction

Base flipping (also known as base eversion) is a type of local DNA motion in which a base group loses the hydrogen bonds with its base pair partner and is everted from the intra-helical to extra-helical position¹. Base flipping was first observed in the DNA/methyltransferase complex X-ray crystal structure^{2,3}. Studies have shown that base flipping is a common strategy for enzymes to read and chemically modify base groups which are otherwise protected by their base pair partner, and/or their own sugar and phosphate groups¹. A variety of these enzymes exist such as methyltransferases, glycosylases, and endonucleases. A number of crystal structures with everted DNA base groups inside the

Correspondence to: Carlos Simmerling.

Supporting Information

Figures of structures with everted base, and sample input file listing atoms in COM restraint groups for CPD definitions. This material is available free of charge via the Internet at pubs.acs.org.

active site of the enzyme have been published^{4–8}. These structures reveal the conformations of the everted base groups, but they provide little insight into the conformational changes that occur during the flipping process, and more importantly, the possible transient role of enzyme functional groups in the facilitation of base flipping. The rate of base flipping can be measured experimentally by methods such as proton exchange^{9,10}. However, studies have shown that this method may overestimate the flipping rate since proton exchange may occur in structures with limited solvent accessibility and thus not require complete base flipping¹¹.

Several computational approaches have been applied to this subject^{11–19}. It is currently necessary to force eversion using restraints in order to model the process during computationally tractable simulations since uncatalyzed base flipping occurs on the millisecond timescale⁹. In one of the earliest studies, Keepers et. al. used a distance restraint between the N1(pyrimidine) and N3(purine) to force the base pair to break^{20,21}. However, the distance restraint can not specify which of the two bases to flip, or distinguish between major or minor groove flipping pathways. Inspired by the correlation between the base opening angle and the ζ dihedral angle seen in the crystal structure, Chen et. al. applied restraint forces on the ζ angle and glycosidic angles of the target nucleotide to force base pair opening²². This procedure assumes that only two dihedral angles of the targeted nucleotide are responsible for base flipping, which may not be generally true. It has been shown that using this method can generate artificial conformations¹². More promising approaches have employed more sophisticated reaction coordinates; these have been applied by Lavery and coworkers using internal coordinates^{13–16} and MacKerell and coworkers using a center of mass pseudodihedral angle^{11,17–19} (the latter is hereafter referred to as the CPD angle). The results from these two approaches are in reasonable agreement. Due to the nature of its definition using a standard dihedral angle with points defined by center of mass groups, MacKerell's pseudodihedral angle method is relatively easier to implement in current molecular dynamics simulation algorithms. An excellent recent application of this method to cytosine 5-methyltransferase from HhaI suggests that the enzyme shifts the equilibrium to the flipped state by destabilizing the DNA duplex and stabilizing the everted conformation¹⁷.

In the present study we first tested the CPD definition to study base flipping using the Amber simulation package and then improved the CPD definition²³. Our results show that MacKerell's CPD definition can have several potential disadvantages in spite of its multiple strengths. Using the traditional CPD definition we find that there are large energy fluctuations and occasional simulation instability when the base is everted. Another disadvantage is that there is not a unique mapping between the CPD angle and the extent of eversion; we observed that significantly different structures can have comparable CPD angles when using the previous definition. Based on our MD simulations of spontaneous base pair formation in unrestrained simulation, we identified several reasons for the weaknesses described above. We thus modified the CPD definition to employ two separate definitions of different groups of atoms to define the pseudo-dihedral angle (hereafter referred to as CPDa and CPDb); we find that these better represent base flipping. Simulations using the CPDa/b reaction coordinates were able to give a more reliable representation of the base flipping pathway with improved correlation between the reaction coordinate and the extent of eversion, along with improved simulation stability. The two new

CPDa/b angles were applied to calculate the free energy profile of base flipping for guanine in a DNA duplex. The free energy barrier of the base flipping is comparable with previous studies. The relationship between the specific definition of the reaction coordinate and the overall shape of the resulting free energy profile are discussed.

2. Methods

2.1 System preparation

The initial structure was standard B-form duplex DNA built using the NUCGEN program in the Amber simulation package²³. The DNA sequence used is shown in table 1 below. The targeted central base pair for flipping was C10:G24, and G24 was the base group (hereafter referred as target base group) for base flipping.

The initial coordinate and parameter files were created using the LEAP module of Amber based on the structure generated by NUCGEN, solvated in truncated octahedron boxes with a minimum 8 Å buffer between the box edge and the nearest solute atom. The TIP3P water model²⁴ was used to explicitly represent water molecules. The deoxyribonucleic acid parameters employed ff99^{25,26}, with the parmBSC0 modified α/γ torsional terms²⁷. These coordinates were used for all simulations.

2.2 Molecular Dynamics simulations

All molecular dynamics simulations were carried out with the SANDER module in Amber²³. Following the procedure used in previous studies^{28,29}, the solvated systems were minimized and equilibrated in three steps: (i) 50 picoseconds MD simulation³⁰ with DNA atoms constrained and movement allowed only for water; (ii) five 1000-step cycles of minimization, in which the positional restraints on the DNA were gradually decreased; (iii) four cycles of 5000 steps MD simulation with decreasing restraints on the DNA. A final 500 ps of MD were performed without restraints. The resulting structures were used in umbrella sampling simulations.

SHAKE³¹ was used to constrain bonds involving hydrogen atoms. The non-bonded cutoff was 8 Å. The particle mesh Ewald method^{32,33} was used to calculate long-range electrostatics. Constant pressure (1 atm) and temperature (330 K, slightly elevated to improve sampling) were maintained by the weak coupling algorithm with a coupling constant of 1 ps³⁴. A 0.002 ps time step was used.

2.3 Structural analysis

To obtain unbiased structural analyses on the three types of reaction coordinates presented in this work, we generated an unrestrained MD simulation starting from a structure containing an everted target base, which then spontaneously reinserted into the duplex during the simulation. In this simulation, the same sequence of DNA duplex and the simulation conditions described in the previous sections were used. The root mean square deviation (RMSD) of all heavy atoms in the central 3 base pairs was calculated, with the reference structure of a standard B-form duplex DNA. The glycosidic angle of the G24 nucleotide was

also measured, using the dihedral angle of O4'-C1'-N9-C4. The distance between the central base pair is represented by the distance between N1 of G24 and N3 of C10.

2.4 The definition of the base opening dihedral angle

Umbrella sampling³⁵⁻³⁸ was used to calculate the potential of mean force (PMF) as a function of our new center of mass pseudo-dihedral angles CPDa/b. The definitions of two proposed variations on this new flipping metric, CPDa and CPDb, are shown in Figure 1. The Sander module of Amber9 was modified to support these restraints. The four points for each dihedral form two triangular planes which share one side defined between P2 and P3. The base opening angle is defined by the angle between these two planes. Changes from the definition of MacKerell et al. involve the use of both flanking base pairs for the P1 center of mass (as opposed to only one flanking pair), and using either the sugars or phosphates flanking the flipping base as points P2 and P3. Point P4 was defined using only the 5-member ring of the purine in order to remove the influence of glycosidic rotation on the flipping angle, which occurs if the entire purine base is included in point P4 (data not shown).

2.5 Umbrella sampling and potential of mean force calculations

The procedure for umbrella sampling was adapted from previous studies¹¹. Starting from the standard B-form conformation, the initial structure of each window was generated by 0.5 ps simulation with a restraint force constant of 10,000 kcal/(mol×radian²) in a serial fashion, which used the previous window's last structure as the starting structure of the current window. Each window was separated by 5 degrees from the flanking windows. After the initial structures were generated, 500 ps simulations using the same definition of restraint and a 1,000 kcal/(mol×radian²) restraint constant were carried out for the sampling. The eversion angle and energy data were recorded each time step. The other parameters of these simulations were the same as the standard MD simulations. The resulting PMF was obtained by WHAM analysis³⁶⁻³⁸ of the data using a program provided by Alan Grossfield (freely available at <http://dasher.wustl.edu/alan>).

3. Results and discussion

3.1 Simulation of spontaneous base pair formation

Umbrella sampling simulations provide the PMF along the chosen reaction coordinate for a system. To generate an accurate free energy profile along a conformational change of interest, the reaction coordinate should be able to represent conformational change faithfully. To evaluate the extent to which various reaction coordinates can represent the process of a base flipping event, we generated an unrestrained trajectory for a duplex in explicit water where a base pair spontaneously formed from the everted position, since it is a faster event than base opening. The starting structure was obtained from an umbrella sampling simulation using the CPD definition at -140°. A 10-ns unrestrained MD simulation was carried out. From the original everted position, the guanine spontaneously returned to the duplex and reformed the Watson-Crick base pair with the cytosine in less than 5 ns. Several snapshots taken from the trajectory are shown in Figure 2. In the starting structure (0 ps), the everted base (G24) is completely outside of the duplex. At ~1 ns, it moved closer to the

major groove and the purine ring adopted a conformation nearly perpendicular to the other, stacked base groups. This conformation was stable until ~3 ns, during which the purine ring attempted to reinsert into the duplex, but was unsuccessful since the G24 was still in the *syn* conformation. It returned back to its previous everted position at ~4 ns, adopted an *anti* conformation at ~4.2 ns, and then successfully reinserted into the duplex at ~4.3 ns. The newly formed base pair was stable for the remainder of the 10 ns simulation (only 5 ns is shown in Figure 2).

The observation of spontaneous base pair formation provides an excellent data set for evaluation of the various parameters in the base eversion restraints that will be used for umbrella sampling. In particular, any measure of base flipping should reproduce the observation that the first attempt at reinsertion by the base was unsuccessful, after which it moved back out of the major groove and then successfully inserts. We calculated several properties of the DNA duplex along this trajectory, such as the distance between the forming base pair (represented by the distance between N1 of G24 and N3 of C10), the glycosidic angle of the flipped G24 base, and the RMSD value of the central three base pairs relative to the standard B-form DNA. The results are shown in Figure 3.

At the beginning of the simulation, the distance between the central two bases was ~ 15 Å. The flipped base G24 was in a *syn* conformation, with the glycosidic angle at ~ 55°. The RMSD value of the central 3 base pairs was ~ 5 Å compared to standard B-form DNA. The distance decreased to about 3 Å at ~4.3 ns; further analysis confirmed that this was accompanied by the formation of all three Watson-Crick hydrogen bonds. The reformed hydrogen bonds were stable for the rest of the simulation. Also at 4.3 ns, the glycosidic angle changed to ~ -90°, falling within the *anti* range, and the RMSD value decreased to ~ 1.5 Å, indicating that the structure was highly similar to a canonical B-form DNA.

3.2 Evaluation of alternate base flipping reaction coordinates

The unrestrained MD trajectory exhibited a pathway of spontaneous base pair reforming, which provides an excellent data set for evaluation of the CPD reaction coordinate using COM groups as defined by MacKerell et. al., and comparison to the modified approach with different selections for the COM groups (Figure 1). CPD denotes the base opening angle calculated using MacKerell's et. al.'s center of mass pseudo-dihedral angle (Figure 1A). CPDa and CPDb are the new center of mass pseudo-dihedral angles (Figure 1, B and C). During the analysis, we have found that in the original CPD definition, the center of mass of the flipped base and the next two centers of mass can become collinear (Figure 4, structure image shown in Figure S1). Therefore we have also measured angles defined by these three neighboring COM positions for each CPD definition using the angle defined by points 2–3–4 shown in Figure 1a. Instability in the dihedral calculation should be expected if this angle approaches 0° or 180°.

From Figure 4 we can see that there are several potential disadvantages using the COM groups as originally defined by the original CPD angle. The first disadvantage is that the reaction coordinate does not have a one-to-one correlation between the measured and actual extent of eversion. For example, CPD-dihedral angle values sampled for everted conformations (between 2100 ps and 3100 ps) are comparable to those sampled after the

base pair has formed at 4300 ps. This means two different points on the base everting pathway will have the same value of the reaction coordinate. This is caused by the definition of the CPD angle; the last three points in the CPD definition (P2, P3, and P4, see Figure 1a) can become collinear in everted conformations, resulting in numerical instability. According to the original CPD definition, negative values denote flipping along a major groove pathway, and positive values indicate the minor groove pathway. While the sign of a particular flipping direction (major/minor) is arbitrary and depends on whether the CPD is defined from the 3' or 5' side of the flipping base, the data should be consistent once a definition of CPD is chosen. However, although the trajectory was visually confirmed to sample only the major groove pathway, the CPD during the simulation adopted both negative and positive values, suggesting that the sign is not a reliable indicator of the flipping directions.

By using the new definitions (CPDa and CPDb in Figure 4), the reaction coordinate values and the position of the base group have an improved correlation. The CPDa and CPDb angles gradually reduced from $\sim 180^\circ$ (extrahelical) at the beginning of the simulation to 50° at 2300 ps. The flipped base was close to its intrahelical position at 2300 ps, except that the base was in its *syn* conformation, not in the *anti* conformation required for proper Watson-Crick pairing (See the χ_{24} in Figure 3). Steric hindrance with the phosphodiester backbone prevents rotation about the glycosidic bond in this position, thus the base once again moved out of the major groove, with the CPDa and CPDb correctly reflecting this change, with values increasing between 3600 ps and 4200 ps. The extra-helical base then rotated to an *anti* conformation and subsequently reinserted, restoring the Watson-Crick pair. This is represented with the low (~ 0) and steady values of the CPDa and CPDb flipping angles in Figure 4; we note that with the CPDa and CPDb definitions the intrahelical values (near 0) were not seen for any of the everted conformations. This is in contrast to the “intrahelical” CPD values observed at multiple points prior to the actual reinsertion event.

The second disadvantage of the original pseudodihedral definition is that one of the two angles (CA in Figure 4) connecting the four centers of mass can adopt values close to 0° or 180° when the base is extrahelical. From Figure 4 we can see that angle CA is very close to 180° before the base pair is reformed at about 4000 ps (the distance in panel 1 of Figure 4 can be used as an indicator of the base pair reforming.). After the base pair formed (the distance becomes a steady line at about $3\text{\AA}$), the CA angle adopted values near 90° . The dihedral angles are defined by four points, where each three consecutive points define a plane. The dihedral angle is the angle between these two planes. When the last three consecutive points are close to being linear (0° or 180°), a slight change of the position of the fourth point can greatly change the definition of the second plane, which results in large fluctuations in the dihedral angle (the second panel in Figure 4) and resulting forces. This caused unpredictable instabilities in simulations with everted bases (data not shown); CPD dihedral angle differences of $\sim 40^\circ$ were observed for nearly identical structures sampled during short time spans (Figure S1). In the new definitions, both angles (PA, SA in Figure 4) had values well away from 0° and 180° during the entire profile, resulting in improved numerical stability of the dihedral angle.

The third disadvantage of the old definition is less explicit. The definition of CPD is not symmetric (Figure 1) and the choice of P1 and P2 dihedral points as being either the 3' or 5' side of the flipping base is arbitrary. Due to the asymmetric structure of the DNA duplex, a free energy profile calculated using the two points from 3' side differs from that obtained with restraints for the dihedral points defined on the 5' side. By using the new definition, the 5' and 3' sides are both included in a single calculation, with the resulting PMF being less ambiguous.

3.3 The free energy profiles calculated using the new definitions

The free energy profiles for base eversion using our two new CPDa/b dihedral angle definitions in umbrella sampling have been calculated (Figure 5). The upper panel shows the PMF profile using the centers of mass of the deoxyribose rings groups as dihedral points P2 and P3 (CPDa, Figure 1b). The lower panel shows the results using the phosphates as P2 and P3 (CPDb, Figure 1c). To estimate the convergence of the calculation, we also calculated the free energy profile using the second half of the data and generated error bars using the difference between the two results.

From Figure 5 we can see that there are certain similarities and differences between the two energy profiles. Both profiles can be divided into two regions: the “basin” and the “plateau” regions. The basin region is near the energy minimum and has lower free energy and steeper slope. The plateau region reflects everted bases, and is further from the minimum with high free energy value and less energetic dependence on angle. The basin region for the profile using the phosphate groups (CPDb) was between -45° and 45° . The basin using the sugar group for the PMF reaction coordinate (CPDa) is significantly broader, ranging from -120° to 120° . In both cases, the barrier of the PMFs (after the basin region ends) at the major groove pathway were lower than those at the minor groove pathway. The values for using the sugar rings were about 13 kcal/mol for the major groove and 18 kcal/mol for the minor groove. The values for using phosphate groups were about 12 kcal/mol for the major groove, and 15 kcal/mol for the minor groove. These results agree reasonably with calculations by Banavali et. al. showing that the energy barriers for G flipping are 18.7 and 21.3 kcal/mol for major and minor groove respectively¹¹. The positions of the minima and height of the energy barriers of Banavali's and our studies are similar but do not exactly match. One possible reason is that the sequence contexts are different between these two studies. However, both experimental and theoretical studies have suggested that base opening rates have little dependence on sequence context^{9,39}. The differences may also be due to the influence of the reaction coordinate definition on the PMF details, or enforcement of periodicity in the free energy calculation, or the difference between CHARMM and Amber force fields. The present results are also comparable to the Amber results of Priyakumar et al.¹¹, though that study used an older version of the Amber DNA force field than used here.

Although the free energy barriers were similar in the two free energy profiles in Figure 5, the widths of the basin regions were significantly different. The basin is much wider for the CPDa definition. To understand why the free energy profiles are different, we calculated the correlations between CPDa and CPDb definitions for structures sampled in the two umbrella sampling runs, which are shown in Figure 6. Data from both simulations has a high

correlation between CPDa and CPDb values. However, both correlation figures are not straight lines. There is a flat phase near 0° , which corresponds to the intrahelical conformation of the base. In other words, the intrahelical space is wider using CPDa as reaction coordinates than using CPDb, even for the same structure sets. This is due to the different the geometry of the two definitions in the intrahelical region. In Figure 4 we can see that CPDa is has obtuse angle SA, and PA of CPDb is $\sim 60^\circ$ when the base is in its intrahelical conformation. Since it corresponds to the intrahelical conformation, the basin region is wider in CPDa space.

We further analyzed the correlation between flipping angle and base pair distance to investigate the properties of these two reaction coordinate definitions (Figure 7). Figure 7a shows the structures sampled in the umbrella sampling simulation using the CPDa restraints. Figure 7b shows the structure sampled in the umbrella sampling simulation using CPDb restraints. For all structures in each simulation, we calculated the distance between the flipping base and its partner, as well as flipping angles measured using both dihedral definitions. As we observed with the comparison of the two dihedral angles in Figure 6, data from simulations performed with either definition as the restraint are consistent. In both cases the base pair distance was about 3 Å and stable for the region around a base opening angle of 0° . In both simulations, the region of close contact between the bases covers a significantly larger range when using CPDa as reaction coordinate than when using CPDb as reaction coordinate, even when they were applied to the same set of structures. This confirms that CPDb has a more sensitive to the true extent of base opening. We can also see that the structures sampled using CPDb (Figure 7b) are more similar among windows than the ones sampled using CPDa (Figure 7A). The structures sampled using CPDa (Figure 7a) seems more poorly converged. The reason CPDa and CPDb behave differently is due to the number of internal coordinates encompassed in the two definitions. CPDb includes two phosphate groups and one sugar ring between P2 and P3 while CPDa includes 2 phosphate groups and 3 sugar rings (see Figure 1). Since the CPDa definition has a more complex conformation space, it is more sensitive to structural fluctuations in the backbone.

4. Conclusion

Base flipping is an important event, and computational tools have been shown to be essential in studying processes such as an enzymatic role in flipping. Starting from a flipped conformation, we have generated a fully unrestrained MD trajectory in which an everted guanine base spontaneously returned to its intrahelical conformation and reforming its Watson-Crick pair with the cytosine partner. This trajectory was used to evaluate a previously proposed pseudo-dihedral angle and how well it describes the extent of eversion. We found several disadvantages in the definition, including the potential for numerical instability, and used the data to propose two modified pseudodihedral definitions which can successfully avoid the observed disadvantages. The free energy profiles of the base flipping using the new definitions have been calculated, and the results reasonably agree with previously published results. We also compared the two modified definitions, the one using the center of mass of the phosphate groups has a tighter correlation with base opening angle, therefore it is a better representation for base flipping. The reaction coordinate using the center of mass of the sugar groups has a larger conformation space and appeared to be more

difficult to generate well-converged data. In closing, we remind the reader that using any restraint to impose a reaction coordinate may introduce artifacts in the data as compared to fully unrestrained systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for this project was provided by NIH GM6167803 (CLS), CA17395 (APG & CLS), NSF 0549370 (CB and AJC) and supercomputer resources through the National Computational Science Alliance grant MCA02N028 (CLS).

References

1. Roberts RJ, Cheng XD. Base flipping. Annual Review of Biochemistry. 1998; 67:181.
2. Klimasauskas S, Kumar S, Roberts RJ, Cheng XD. HhaI Methyltransferase Flips Its Target Base out of the DNA Helix. Cell. 1994; 76:357–369. [PubMed: 8293469]
3. Reinisch KM, Chen L, Verdine GL, Lipscomb WN. The Crystal-Structure of HaeIII Methyltransferase Covalently Complexed to DNA - an Extrahelical Cytosine and Rearranged Base-Pairing. Cell. 1995; 82:143–153. [PubMed: 7606780]
4. Sugahara M, Mikawa T, Kumasaka T, Yamamoto M, Kato R, Fukuyama K, Inoue Y, Kuramitsu S. Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. Embo J. 2000; 19:3857–69. [PubMed: 10921868]
5. Fromme JC, Verdine GL. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. Nat Struct Biol. 2002; 9:544–52. [PubMed: 12055620]
6. Fromme JC, Verdine GL. DNA lesion recognition by the bacterial repair enzyme MutM. J Biol Chem. 2003; 278:51543–8. [PubMed: 14525999]
7. Gilboa R, Zharkov DO, Golan G, Fernandes AS, Gerchman SE, Matz E, Kycia JH, Grollman AP, Shoham G. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. J Biol Chem. 2002; 277:19811–6. [PubMed: 11912217]
8. Serre L, Pereira de Jesus K, Boiteux S, Zelwer C, Castaing B. Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. Embo J. 2002; 21:2854–65. [PubMed: 12065399]
9. Leroy JL, Kochoyan M, Huynhdinh T, Gueron M. Characterization of Base-Pair Opening in Deoxynucleotide Duplexes Using Catalyzed Exchange of the Imino Proton. Journal of Molecular Biology. 1988; 200:223–238. [PubMed: 2836594]
10. Coman D, Russu IM. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. Biophys J. 2005; 89:3285–3292. [PubMed: 16126830]
11. Banavali NK, MacKerell AD. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. Journal of Molecular Biology. 2002; 319:141–160. [PubMed: 12051942]
12. Priyakumar UD, MacKerell AD. Computational approaches for investigating base flipping in oligonucleotides. Chemical Reviews. 2006; 106:489–505. [PubMed: 16464016]
13. Giudice E, Varnai P, Lavery R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations (vol 31, pg 1434, 2003). Nucleic Acids Research. 2003; 31:2703–2703.
14. Giudice E, Varnai P, Lavery R. Energetic and conformational aspects of A : T base-pair opening within the DNA double helix. Chemphyschem. 2001; 2:673. [PubMed: 23686902]
15. Giudice E, Lavery R. Nucleic acid base pair dynamics: The impact of sequence and structure using free-energy calculations. Journal of the American Chemical Society. 2003; 125:4998–4999. [PubMed: 12708848]

16. Varnai P, Lavery R. Base flipping in DNA: Pathways and energetics studied with molecular dynamic simulations. *Journal of the American Chemical Society*. 2002; 124:7272–7273. [PubMed: 12071727]
17. Huang N, Banavali NK, MacKerell AD. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:68–73. [PubMed: 12506195]
18. Huang N, MacKerell AD. Specificity in protein-DNA interactions: Energetic recognition by the (cytosine-C5)-methyltransferase from HhaI. *Journal of Molecular Biology*. 2005; 345:265–274. [PubMed: 15571720]
19. Priyakumar UD, MacKerell AD. Base flipping in a GCGC containing DNA dodecamer: A comparative study of the performance of the nucleic acid force fields, CHARMM, AMBER, and BMS. *Journal of Chemical Theory and Computation*. 2006; 2:187–200. [PubMed: 26626393]
20. Keepers J, Kollman PA, James TL. Molecular mechanical studies of base-pair opening in d(CGCGC):d(GCGCG), dG5 . dC5, d(TATAT):d(ATATA), and dA5 . dT5 in the B and Z forms of DNA. *Biopolymers*. 1984; 23:2499–511. [PubMed: 6518264]
21. Keepers JW, Kollman PA, Weiner PK, James TL. Molecular mechanical studies of DNA flexibility: coupled backbone torsion angles and base-pair openings. *Proc Natl Acad Sci U S A*. 1982; 79:5537–41. [PubMed: 6957879]
22. Chen YZ, Mohan V, Griffey RH. The opening of a single base without perturbations of neighboring nucleotides: a study on crystal B-DNA duplex d(CGCGAATTCGCG)2. *J Biomol Struct Dyn*. 1998; 15:765–77. [PubMed: 9514252]
23. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 2005; 26:1668–1688. [PubMed: 16200636]
24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*. 1983; 79:926–935.
25. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society*. 1995; 117:5179–5197.
26. Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*. 2000; 21:1049–1074.
27. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*. 2007; 92:3817–29. [PubMed: 17351000]
28. Song K, Hornak V, de Los Santos C, Grollman AP, Simmerling C. Computational analysis of the mode of binding of 8-oxoguanine to formamidopyrimidine-DNA glycosylase. *Biochemistry*. 2006; 45:10886–94. [PubMed: 16953574]
29. Song K, Hornak V, De Los Santos C, Grollman AP, Simmerling C. Molecular mechanics parameters for the FapydG DNA lesion. *Journal of Computational Chemistry*. 2007
30. Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE 3rd, Dixit SB, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Seibert E, Sklenar H, Stoll G, Thayer KM, Varnai P, Young MA. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J*. 2004; 87:3799–813. [PubMed: 15326025]
31. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics*. 1977; 23:327–341.
32. Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*. 1993; 98:10089–10092.
33. Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA. Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, Rna, and Proteins. *Journal of the American Chemical Society*. 1995; 117:4193–4194.

34. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics*. 1984; 81:3684–3690.
35. Kottalam J, Case DA. Dynamics of Ligand Escape from the Heme Pocket of Myoglobin. *Journal of the American Chemical Society*. 1988; 110:7690–7697.
36. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *Journal of Computational Chemistry*. 1992; 13:1011–1021.
37. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *Journal of Computational Chemistry*. 1995; 16:1339–1350.
38. Roux B. The Calculation of the Potential of Mean Force Using Computer-Simulations. *Computer Physics Communications*. 1995; 91:275–282.
39. Krueger A, Protozanova E, Frank-Kamenetskii MD. Sequence-dependent base pair opening in DNA double helix. *Biophys J*. 2006; 90:3091–9. [PubMed: 16500982]

Figure 1.

Definitions of various reaction coordinates for base eversion. A). CPD: MacKerell et. al.'s original COM pseudo-dihedral angle definition. B). CPDa: the modified COM pseudo-dihedral (CPDa) angle definition, in which p1 is defined by the mass center of the two flanking base pairs, p2 and p3 are defined by the flanking sugar groups, and p4 is defined by the five member ring of the flipping purine (or the entire six-membered ring for a flipping pyrimidine). C). CPDb: a similar definition as CDPa, but using the phosphate groups for p2 and p3. The dotted lines show the two planes which define the pseudo-dihedral angles.

Figure 2. Snapshots taken from the base pair reforming trajectory viewed from the major groove. For clarity, water, hydrogen atoms and DNA outside the central three base pairs are not shown. The simulated structures are colored by atom type. The structure in gray indicates the same duplex in a standard B-form conformation for reference. The time sequence is described in the text.

Figure 3.

Data from the unrestrained trajectory with spontaneous base pair formation. The upper panel shows the heavy atom to heavy atom hydrogen bonding distance between atom N1 of residue G24 and atom N3 of residue C10 between the bases in the new base pair. The middle panel shows the glycosidic angle of the flipped G24 nucleotide. The lower panel shows the RMSD of all atoms in the central 3 base pairs as compared to standard B-form DNA.

Figure 4.

The evaluation of the reaction coordinates defined by MacKerell et. al. and the new definitions. The first panel is the distance between the two bases in the base pair being formed which is heavy atom to heavy atom hydrogen bonding distance between residue G24 atom N1 and residue C10 atom N3. The pseudodihedral angles are defined as the angles between two planes (see Figure 1). CPD is the original dihedral reaction coordinate. CPDa is the pseudo-dihedral angle using sugar groups, CPDb is the angle using phosphate groups. Angles between points 2, 3 and 4 for CPD, CPDa and CPDb are shown as CA, SA (sugar angle) and PA (phosphate angle), respectively. Instability in the dihedral results when these angles adopt values very near 0° or 180° .

Figure 5.

Free energy profiles using the two new pseudodihedral eversion definitions. The x axis is the pseudo-dihedral angle. Panel A shows the result using the center of mass of the sugar rings as points P2 and P3 (CPDa). Panel B shows the result using the center of mass of the phosphates as dihedral points P2 and P3 (CPDb). Positive/negative values reflect flipping into major/minor grooves respectively. The solid line shows the free energy profile calculated using the last 400 ps of total 500 ps per window. The error bar shows the difference of the results calculated using the last 400 ps data and the last 200 ps data.

Figure 6.

The correlation of the two CPDa/b definitions. Panel A shows the structures sampled using the CPDa restraint. Panel B shows the structures sampled using the CPDb restraint. The x axis shows the post-processing results using CPDa, while data for CPDb is are on the y axis. Regardless of the restraint used to generate the structures, CPDa is more sensitive than CPDb to changes in the region near 0° .

Figure 7.

The correlation between the distance of the two bases (distance between atom N1 of residue G24 and atom N3 of residue C10) and the CPDa/b base opening angles. The upper figure shows the structures sampled in the umbrella sampling using CPDa as the reaction coordinate. The lower figure shows the structures sampled in the umbrella sampling using CPDb as the reaction coordinate. For each simulation, the data shown in red is the post-processing result using the CPDa definition, and the one in black is the result using CPDb. Both plots indicate that the CPDa definition has a broader range of values for base-paired structures as compared to CPDb, independent of which restraint was used to generate the flipping trajectory.

Table 1

Sequence of the duplex DNA used in our simulations. The base pair C10:G24 is the targeted central base pair, and G24 is the target base for flipping.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
5'	A	G	G	T	A	G	A	T	C	C	G	G	A	C	G	C	
		C	C	A	T	C	T	A	G	G	C	C	T	G	C	G	T
		32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17