

NMR Imino Proton Exchange Experiments on Duplex DNA Primarily Monitor the Opening of Purine Bases

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

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

Received September 19, 2005; E-mail: alex@outerbanks.umaryland.edu

Base opening or flipping is an important biological event, being involved in DNA repair, bacterial restriction systems and transcription, among others.^{1,2} These roles have, in part, motivated a large number of experimental studies of base opening via NMR imino proton exchange.^{3–6} From these experiments lifetimes of Watson–Crick (WC) base paired (i.e. closed) and open states and the equilibrium constant between those states have been determined. This determination is based on a two-state model of opening, where the open states are differentiated from the closed states by their ability to undergo imino proton exchange. While this model appears to be robust, an inherent limitation is that for a given base pair the exact conformational change being monitored is not known, as imino proton exchange can occur whenever the imino proton is accessible to solvent. This accessibility may be achieved via opening of the base that contains the imino proton (i.e., G or T) or its WC partner, whose opening also exposes the imino proton of its partner.⁷ In this communication, we apply computational methods to investigate base opening in a GC and an AT base pair for sequences on which imino proton exchange data is available. Results from these calculations indicate that the base opening studies using NMR imino proton exchange are primarily monitoring the opening of the G and A bases in GC and AT base pairs, respectively.

Typically, GC and AT base pair lifetimes in DNA are observed to be approximately 10–50 and 1–5 ms, respectively, with the open base lifetime on the order of nanoseconds. The equilibrium constant between the base pair open and closed states is on the order of 10^{-6} – 10^{-7} . While these phenomena are primarily dependent on the strength of base pair interaction energies, it has been shown that they are also sequence-context dependent.^{3–6} However, interpretation of the data is complicated by a variety of interactions (e.g., base–base hydrogen bonding and stacking) contributing to the experimental results. To allow for a more detailed structural interpretation of such results a better understanding of the events being monitored is required. This may be achieved via potential of mean force (PMF) calculations based on molecular dynamics (MD) simulations.^{8–10} While this approach has been applied previously to investigate base opening,^{11–13} it has not been performed simultaneously on both GC and AT base pairs with the results being directly compared to experimental data. Such a comparison is required to ensure that the structural events observed in the MD studies are representative of the experimental regimen.

Free energy profiles (i.e., PMFs) corresponding to opening of the underlined A and its WC partner T in CGCGAATTCGCG and the underlined C and its WC partner G in CATGGCCGCATG were determined via umbrella sampling.^{7,8} These two sequences were chosen due to the availability of experimental data enabling direct comparison with the computed results.^{3,4} All MD simulations were carried out using the CHARMM program^{14,15} with the CHARMM27 all atom nucleic acid force field^{16,17} following a previously published protocol.^{7,8} Briefly, 72 conformers were generated for each of the base flipping processes. These structures were immersed in a water

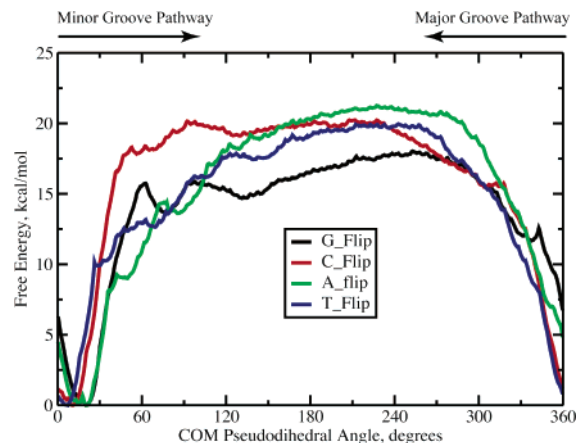


Figure 1. Free energy profiles corresponding to the target G, C, A, and T base flipping processes along the COM pseudodihedral angle coordinate. sphere of radius 35 Å, sodium ions added to maintain electrical neutrality and the solvent molecules equilibrated by a 500-step steepest descent minimization followed by a 20 ps NVT simulation with harmonic restraints on the DNA. For the production run, 72 independent 220-ps MD simulations were performed for each of the base flipping profiles with a harmonic umbrella potential based on a center of mass (COM) pseudodihedral angle (x), where x ranges from 0 to 360° in 5° increments. The pseudodihedral angle was based on the COM of the non-hydrogen atoms in the flipping base, its sugar moiety, the sugar moiety of the adjacent nucleotide on the 3' side, and the adjacent base pair on the 3' side (Figure S1 of Supporting Information). All MD simulations were performed at 288 K in accordance with the experimental conditions. The final 160 ps windows were used for the PMF determination. From the MD simulations probability distributions biased due to the umbrella potential were obtained; weighted histogram analysis method (WHAM) was applied to obtain the unbiased probability distributions from which the free energy surfaces were calculated.^{7,8}

Figure 1 depicts the free energy profiles for opening of the target G, C, A, and T bases along the COM pseudodihedral angle reaction coordinate. The positions of the WC base paired state differ slightly for the four bases ($x = 30^\circ$ for G and A, $x = 7^\circ$ for T, and $x = 10^\circ$ for C) due to differences in the COM of the base atoms used in the definition of x . The free energies of the fully flipped state ($x \approx 195^\circ$) for C, A, and T are similar, around 19.4–20.5 kcal/mol, while the G base opened state lies significantly lower at 16.9 kcal/mol. Along the minor groove pathway, the free energies increase more rapidly compared to the major groove pathway, probably due to the steric hindrance experienced by the flipping base due to its WC partner. However, after the bases move out of the duplex from the WC base paired state by 30° or more, the free energies are typically lower compared to those along the major groove pathway. Free energy profiles corresponding to A and T base flipping indicates the presence of two shallow minima during the early stages of

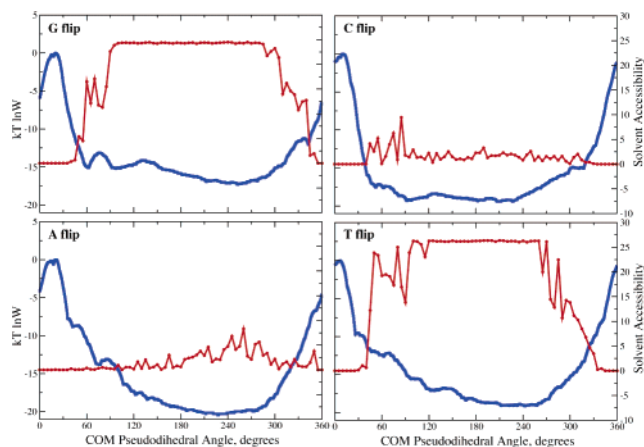


Figure 2. Solvent accessible surface areas in \AA^2 (red) and Boltzmann weighted logarithm of probabilities (blue) as a function of the COM pseudodihedral from the PMF. Regions of the probability distributions designated as open states are listed in Table 1.

flipping along the minor groove pathway. The free energy barrier for the flipping C base is higher than that for the G base, an observation which is different from other studies^{7,12} and appears to be associated with sequence-specific events.

To allow for direct comparison of the calculated PMFs with experimental imino proton exchange data, equilibrium constants for the base open versus closed states were calculated by integrating over the Boltzmann weighted unbiased probability distributions as previously described.⁷ The portions of the PMFs corresponding to the open and closed states were classified on the basis of the solvent accessible surface area (SASA) of the imino proton and the nitrogen to which it is covalently attached for G (for G and C base flipping) and T (for A and T base flipping), by assuming that regions of the PMFs where the SASA is > 0 correspond to open states (Figure 2). The SASA for all four bases increases approximately 45° from the WC base paired state; the smaller change in SASA for the C and A bases is due to the respective G and T bases that contain the imino proton remaining stacked in the DNA helix.

The resulting calculated and experimental equilibrium constants for the AT and GC base pairs are presented in Table 1. The computed equilibrium constants for GC and AT base pair openings are in excellent agreement with the experimental data, validating the applied computational methodology. To understand the contribution of the individual bases to the equilibrium constants, their individual equilibrium constants between the open and closed states are reported in Table 1. As may be seen, the purine bases dominate the overall base pair opening equilibrium constants, which are calculated as the sum of the individual base equilibrium constants. On the basis of this observation, it is predicted that imino proton exchange NMR experiments are monitoring the opening of the purine bases and not the pyrimidine bases, whose contribution to the overall equilibrium constant is lower by close to an order of magnitude or more than those of the purines. This result is consistent with previous data reported for GC opening in a different DNA sequence, that include results from the CHARMM, AMBER 4.1,¹⁸ and BMS¹⁹ force fields.²⁰ Thus, MD simulation results strongly indicate that NMR imino proton exchange experiments are predominately monitoring opening of the purine bases from DNA

Table 1. Calculated Equilibrium Constants for the Base Opening of Individual Bases and the Base Pairs, and the Observed Equilibrium Constants for the GC and AT Base Pair Opening

base pair ^a	equilibrium constants	
	calculated	experimental
GC	1.5×10^{-7}	$(4.8 \pm 0.7) \times 10^{-7}$
AT	4.1×10^{-6}	$(3.7 \pm 0.7) \times 10^{-6}$
Individual Base Contributions		
	COM pseudodihedral ^b	
	angle open range	equilibrium constant
G	$40-350^\circ$	1.5×10^{-7}
C	$45-325^\circ$	7.4×10^{-11}
A	$60-350^\circ$	3.5×10^{-6}
T	$30-340^\circ$	5.8×10^{-7}

^a GC experimental data from ref 3 and AT data from ref 4. ^b The COM pseudodihedral angle ranges corresponding to the selected open states are also given.

duplexes. While this conclusion may be dependent on the DNA sequence, it should facilitate the structural interpretation of the related experimental studies.

Acknowledgment. Appreciation to the NIH (GM51501), the Pittsburgh Supercomputing Center and the NCSA for support, and Dr. Irina Russu for helpful discussion.

Supporting Information Available: Figure defining the COM pseudodihedral angle used for base opening. This information is available free of charge via the Internet at <http://pubs.acs.org/>

References

- Roberts, R. J. *Cell* **1995**, *82*, 9–12.
- Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181–198.
- Dornberger, U.; Leijon, M.; Fritzsche, H. *J. Biol. Chem.* **1999**, *274*, 6957–6962.
- Wärmländer, S.; Sen, A.; Leijon, M. *Biochemistry* **2000**, *39*, 607–615.
- Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383–413.
- Chen, C. J.; Russu, I. M. *Biophys. J.* **2004**, *87*, 2545–2551.
- Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *319*, 141–160.
- Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68–73.
- Huang, N.; MacKerell, A. D., Jr. *Philos. Trans. R. Soc. London, Ser. A* **2004**, *362*, 1439–1460.
- Huang, N.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2005**, *345*, 265–274.
- Varnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272–7273.
- Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434–1443.
- Fuxreiter, M.; Luo, M.; Jedlovsky, P.; Simon, I.; Osman, R. *J. Mol. Biol.* **2002**, *323*, 823–834.
- MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parametrization with an Overview of the Program. 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, pp 271–277.
- Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487–509.
- Priyakumar, U. D.; MacKerell, A. D., Jr. *J. Chem. Theor. Comput.* **2005**, published online November 16, <http://dx.doi.org/10.1021/ct0501957>.

JA056445A