# Supporting Information:

# Base-Flipping Propensities of CpG Sites

Caterina Bianchi[1] and Ronen Zangi[1,2,*]

*1. Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain*

*2. IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

January 30, 2013

* Corresponding author: r.zangi@ikerbasque.org, +34-943018112

In order to examine whether the widths of the grooves play a role in the base-flipping propensities of un-, hemi-, and fully-methylated DNAs (as indicated in Tables 1 and 2 in the manuscript) we calculated these widths and the results are shown in Table S1 below. The width of the minor-groove seems to be largest for the fully-methylated CpG site, whereas, the width of the major-groove is largest for the hemi-methylated site. Thus, this trend exhibits no correlation to, and therefore can not explain, the free energy propensities of flipping-out the cytosine bases.

Table S1: The widths of the major and minor grooves for un-, hemi- and fully-methylated C6pG7 sites. The calculations were performed on the average structures of the trajectories of the alchemical mutations in which both cytosines are intra-helical. The average structures were computed from the covariance matrix analysis in Gromacs[1]. The groove widths were calculated using the Curves+ software[2] at three levels along the CpG sites (6.0, 6.5, and 7.0, in which the first and the last correspond to the C6:G6' and G7:C7' base-pairs, respectively). All values are reported in Å.

| | Major Groove | | | Minor Groove | | |
|---|---|---|---|---|---|---|
| | 6.0 | 6.5 | 7.0 | 6.0 | 6.5 | 7.0 |
| UMe | 14.5 | 13.8 | 13.3 | 7.2 | 6.6 | 5.3 |
| HMe | 16.4 | 16.0 | 16.3 | 6.9 | 6.4 | 6.0 |
| FMe | 14.1 | 14.0 | 14.4 | 8.7 | 7.5 | 8.5 |

Analyses of X-ray DNA structures have shown that the backbone dihedral angels $\epsilon$ and $\zeta$, which are coupled, can exhibits two states, BI and BII. In the more common BI state ($\epsilon - \zeta < 0$) the phosphate group is positioned more symmetrically between the major and the minor-grooves, whereas in the BII state ($\epsilon - \zeta > 0$) it is closer to the minor-groove. Computer simulation studies revealed that C5 methylation of cytosine stabilizes the BI conformation[3]. We calculated the fraction of time the backbone conformation of the two cytosines in the CpG step is in BI conformation (Table S2 below). For the DNA conformation in which both cytosines are intra-helical, the BI state is observed more than 91% of the time and the small differences observed for the different methylation states are probably not significant. However, when one of the cytosines changes its intra-helical state to an extra-helical there is a substantial increase in the stability of the BII conformation. In two cases, a trend in which methylation decreases the stability of the BI state is observed. However, in the other two cases this trend does not hold, eliminating possible correlation with the relative free energy changes for base-flipping.

Table S2: The fraction of time one of the cytosine nucleotides of the CpG site, thus, (m)C6 or (m)C7', is in the BI state. The calculations are performed for unmethylated, hemi-methylated, and fully-methylated CpG sites, in both, normal B-DNA structure (intra-helical state) and in a DNA structure in which one of the cytosines is completely flipped-out of the DNA helix (extra-helical state).

| | Intra-Helical State | | | Extra-Helical State | | |
|---|---|---|---|---|---|---|
| Nucleotide | Un- | Hemi- | Fully- | Un- | Hemi- | Fully- |
| (m)C6 [flipped-in] | 0.99 | 0.97 | 0.98 | 0.99 | 0.50 | 0.01 |
| (m)C7' [flipped-in] | 0.91 | 0.94 | 0.94 | 0.99 | 0.50 | 0.62 |
| (m)C6 [flipped-out] | — | — | — | 0.98 | 0.66 | 0.47 |
| (m)C7' [flipped-out] | — | — | — | 0.50 | 0.82 | 0.01 |

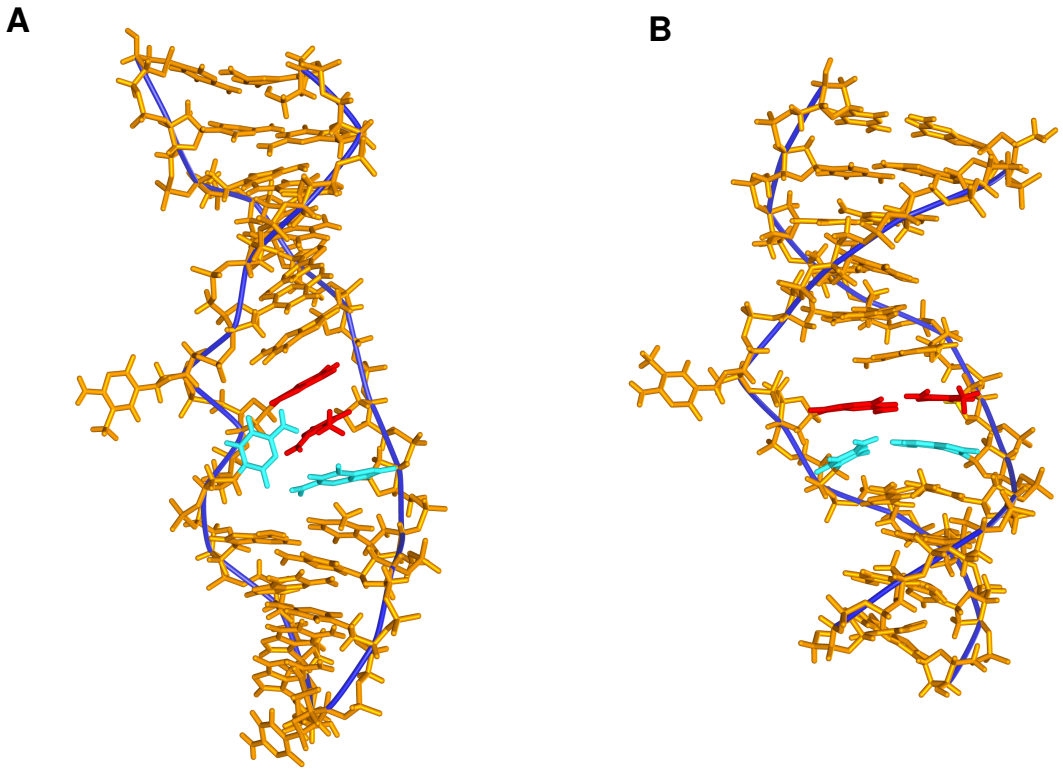


Figure S1: A snapshot of the fully-methylated DNA taken from the alchemical mutation simulations ($\Delta G_4$) showing the deformation of the double-helix and the loss of two base-pairs employing the AMBER99 force-field[4] (a). The employment of the parmbsc0 force-field[5] (b) did not exhibit this behavior. For both structures the base pair G7:C7' is colored in red, whereas the base-pair C8:G8' is colored in cyan. The rest of the DNA duplex and its backbone lines are shown in orange and dark blue, respectively.

## References

[1] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[2] Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+, *Nucleic Acids Res.* **2009**, *37*, 5917–5929.

[3] Temiz, N. A.; Donohue, D. E.; Bacolla, A.; Luke, B. T.; Collins, J. R. The Role of Methylation in the Intrinsic Dynamics of B- and Z-DNA, *PLoS ONE* **2012**, *7*, e35558.

[4] Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comp. Chem.* **2000**, *21*, 1049–1074.

[5] Pérez, A.; March'an, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of $\alpha/\gamma$ Conformers, *Biophys. J.* **2007**, *92*, 3817–3829.