SUPPLEMENTARY INFORMATION

Sequence environment modulates the impact of methylation on torsional rigidity of DNA

Johanna Hörberg[a] and Anna Reymer*[a]

[a] Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg 40530, Sweden.
* To whom correspondence should be addressed. Email: anna.reymer@gu.se

DNA Oligomers

In the study we used six DNA 20-mer oligomers: GG-$(ACGT)_4$-GG, GG-$(CCGC)_4$-GG, GG-$(TCGA)_4$-GG, GG-ACGT-$(AMGT)_2$-ACGT-GG, GG-CCGC-$(CMGC)_2$-CCGC-GG, GG-TCGA-$(TMGA)_2$-TCGA-GG. For every oligomer the torsional restraint was applied to the central eight base pair (bp) region, between the 7th and 15th bp (i.e. 7 bp steps). This was done to avoid potential problems with DNA termini fraying and to focus on twisting deformations. The total twist between the chosen bp was modified, i.e. increased or decreased in steps of 3.5° (corresponding to an average change in twist of 0.5° per bp step. For every following MD simulation, the final structure of the preceding simulation was used as a starting structure to form a so-called cascade umbrella sampling simulation[1], see below. This approach was proven to provide a better convergence. The restrained MD simulation started from fully relaxed structures and worked outward, underwinding and overwinding the restrained regions to a maximum of ±6° with respect to the relaxed averaged twist per bp step for each oligomer.

Molecular Dynamics Simulation Protocol

The DNA oligomers were constructed in standard B-DNA form with the help of nucleic acid modeling program JUMNA[2]. To construct methylated oligomers, the corresponding H5 atom of cytosine was mutated to a methyl group using USCF Chimera program[3]. Unrestrained MD simulations followed by umbrella sampling simulations were performed using GROMACS MD software package, version 5.1[4]. Restrained MD simulations were performed using a structural restraint[5], which controls the torsional state of a DNA molecule, implemented in PLUMED free energy library environment, version 2.2[6]. The torsional restraint uses a force constant of k_{tw} of 0.06 kcal mol^{-1} $degrees^{-2}$ – the smallest value which enables achieving the desired torsional stress. Simulations were carried out using the AMBER all-atom nucleic acid force field Parmbsc1[7], and previously derived parameters for 5-methylcytosine residue[8,9]. Each DNA oligomer was first neutralized with 38 K^+ counterions and solvated with 10 Å layer of TIP3P water[10], contained within a cubic cell under periodic boundary conditions. Additional K^+ and Cl^- ions were then added to achieve a physiological salt concentration of 150 mM. The conformation of each oligomer was initially energy minimized with 5000 steps of steepest descent, followed by a 200 ps simulation at constant volume, while raising the temperature to 300 K. Simulations were then carried out at constant pressure and temperature of 1 atm., and 300 K, using a weak-coupling thermostat[11] with a 0.2 ps coupling constant, and an isotropic Parrinello-Rahman barostat[12] with a 2 ps coupling constant. Simulations used a 2 fs time step; trajectory snapshots were recorded at 1 ps intervals. Bonds involving hydrogen atoms were constrained with the LINCS algorithm[13]. Electrostatic forces were evaluated with particle-mesh Ewald[14] with a real-space cutoff of 10 Å. The van der Waals forces were truncated at 10 Å and long-range corrections were added. Center of mass movement was removed every 0.2 ps to avoid the building up of translational kinetic energy[15]. Weighted Histogram Analysis Method (WHAM) method[16] was

used to correct the probabilities for the impact of the restraining potential during umbrella sampling simulations. We used the version of WHAM implemented in PLUMED, to obtain the corresponding potential of mean force (PMF) with respect to DNA twisting. Following a 300 ns equilibration on each unrestrained oligomer, umbrella sampling simulations were carried out with 300 ns of sampling per umbrella window to allow for counterions equilibration[17]. The initial 50 ns of the sampling time were discarded as equilibration and WHAM procedure was carried on to applied to two equal 125 ns blocks of data to test convergence. For all but ACGT oligomer deviations between the PMF profiles representing the two sampling blocks were negligible. Thus for ACGT oligomer WHAM analysis was applied to 4 50ns-sampling block discarding first 100ns to reach the level of convergence observed for the other sequences. The total simulation time was 3.9μs for each DNA oligomer.

Conformational Analysis of Oligomers

The conformational analysis of the recorded MD trajectories was implemented in several steps: firstly by preprocessing using CPPTRAJ program[18] from AMBERTools 16 software package; then using analysis programs Curves+ and Canal[19], DNA helical parameters, backbone torsional angles, and groove geometry parameters were analyzed for each trajectory snapshot, which provided complete, time-dependent information on the impact of cytosine methylation of DNA response to the imposed torsional stress. MatLab software was used to perform quadratic fitting of the PMF profiles.

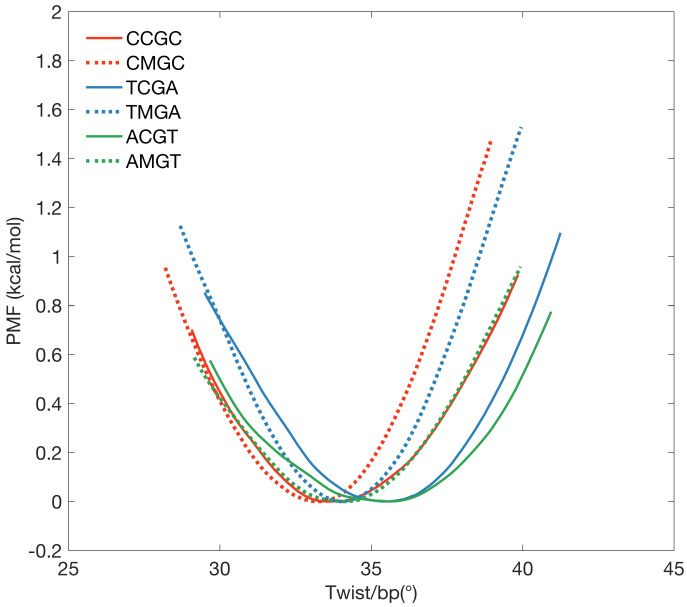

Figure S1. PMF plots showing the change of free energy as a function of average twist per base pair step (bp).

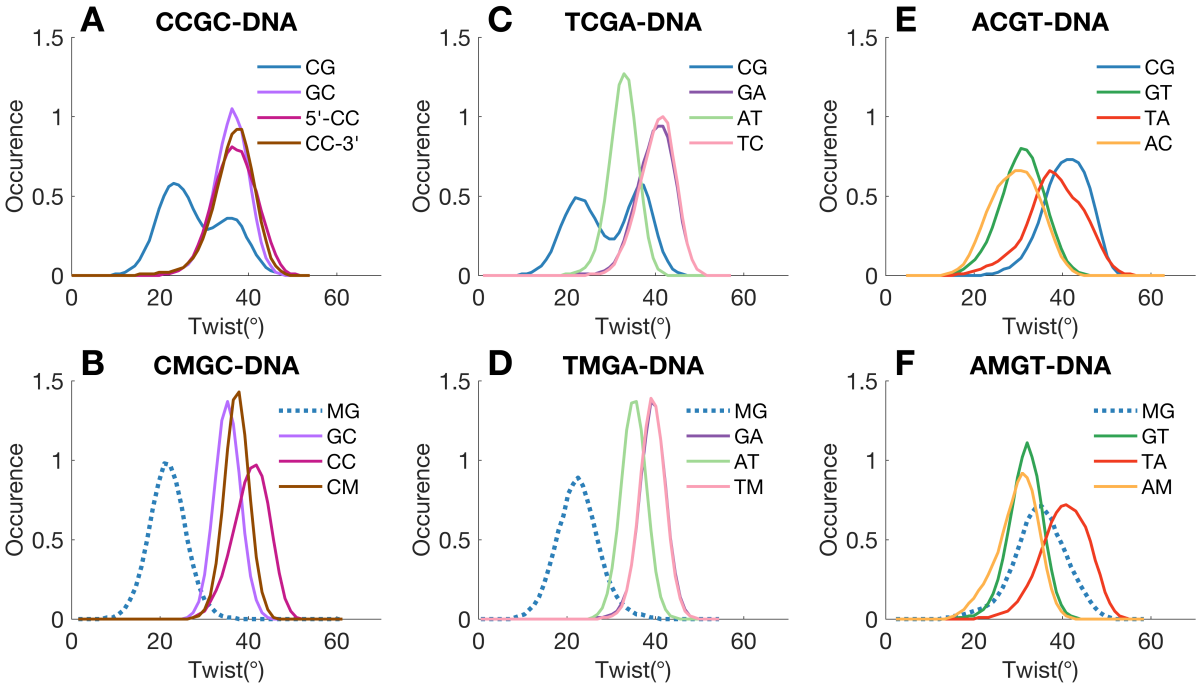

Figure S2. Twist distributions of dinucleotide steps derived from 300 ns unrestrained MD trajectories of the six oligomers.

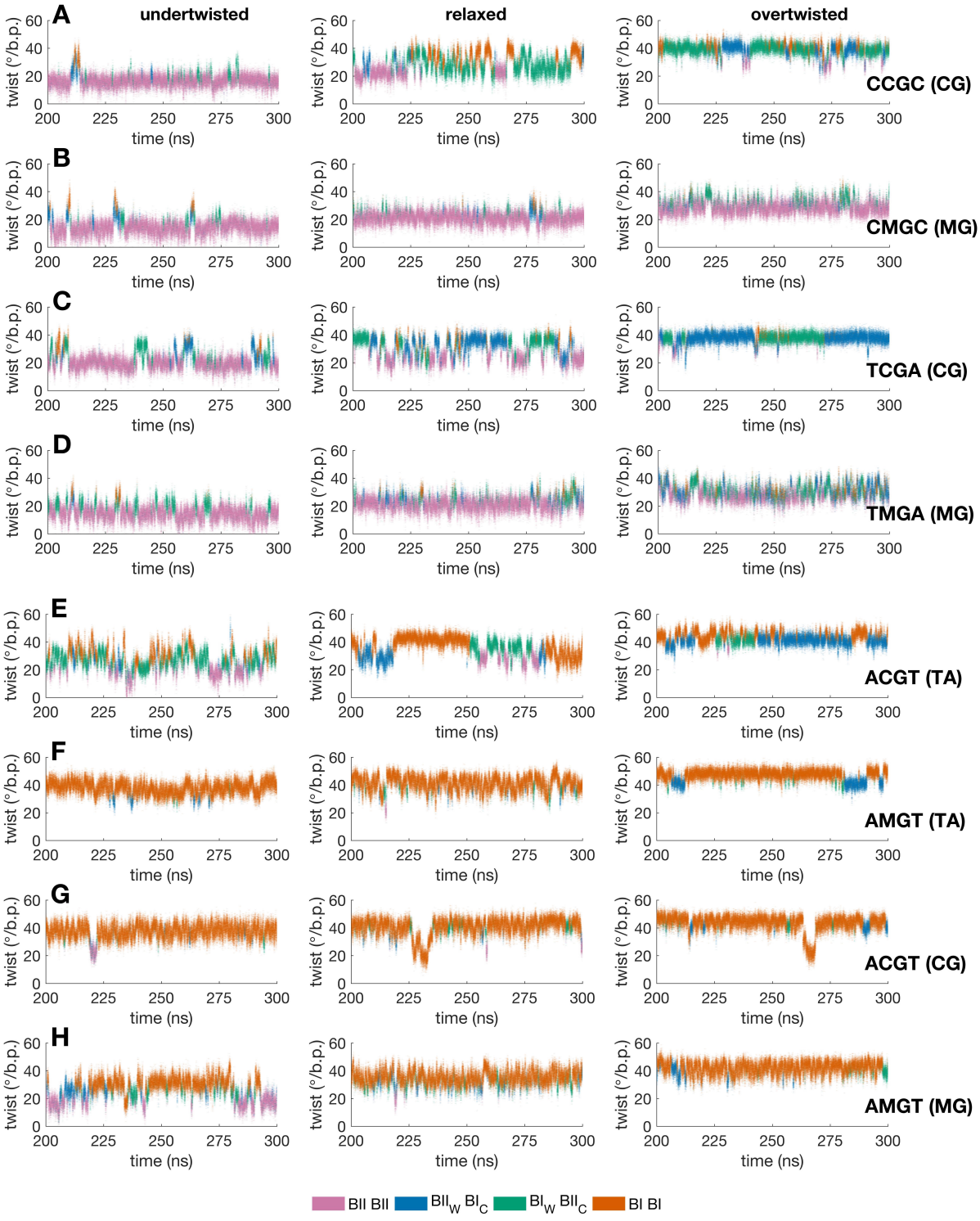

Figure S3. Time evolution of twist for the steps most affected by the imposed twist in each oligomer: CpG/MpG in CCGC/CMGC-DNAs and in TCGA/TMGA-DNAs; TpA and C/MpG in ACGT/AMGT-DNAs. The central panels show the behaviour of these steps in the relaxed oligomer, while the left and right-hand panels show the impact of undertwisting (-5.0°) or overtwisting (+5.0°) with respect to the average base pair step twist of the restrained segments. The time series are coloured as a function of the BI/BII state of the 3'-flanking phosphate junctions for the Watson (w) and Crick (c) strands. The plots show the last 100 ns of the corresponding simulations,

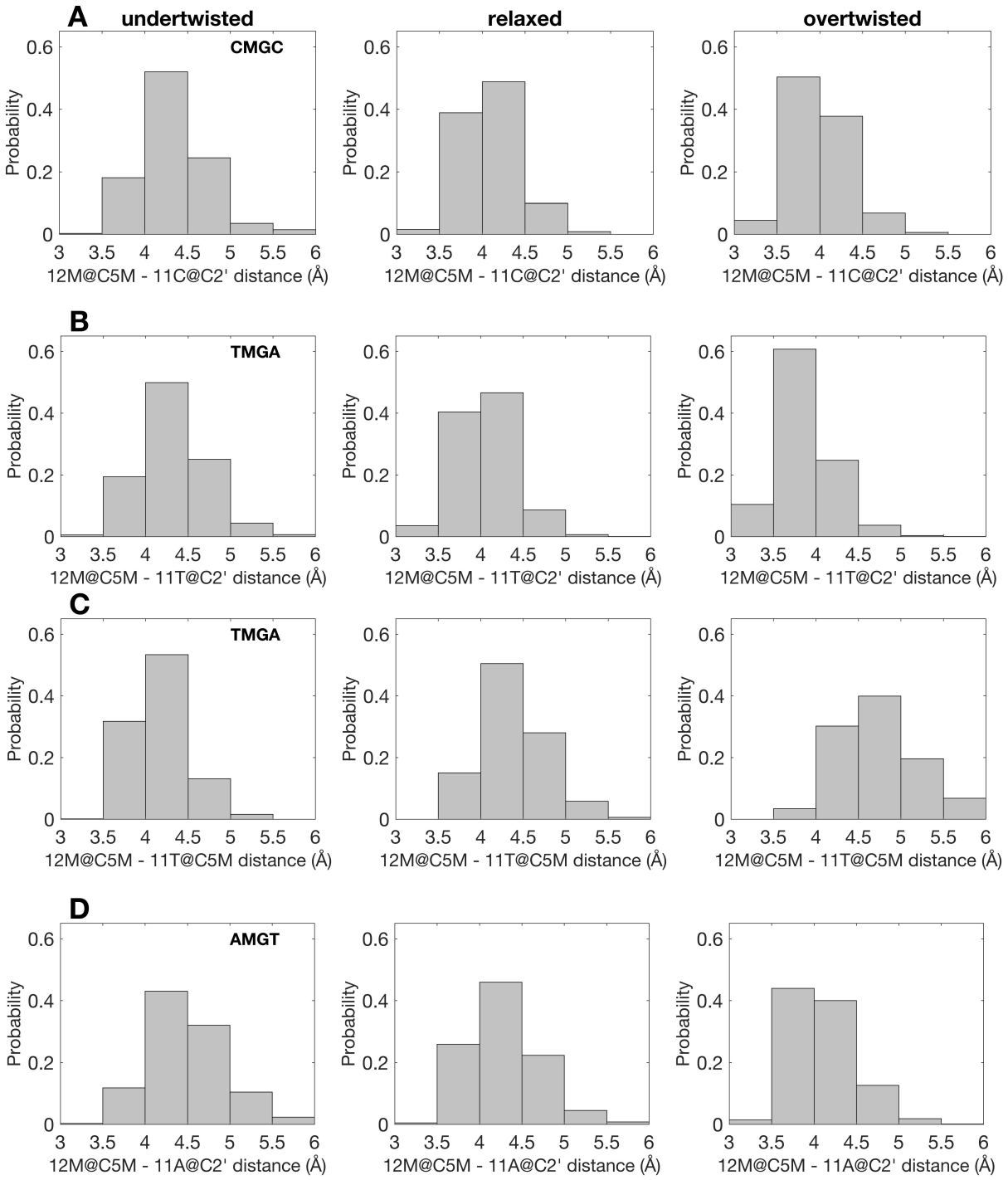

Figure S4. Distance distributions, highlighting potential steric clashes between $5C_M$ atom of 5-methylcytosine and C2' atom of the 5'-flacking nucleotide in CMCG-DNA (A), in TMGA-DNA (B), and in AMGT-DNA (D); between $5C_M$ atom of 5-methylcytosine and $5C_M$ atom of 5'-thymine in TMGA (C). Symbol "@" in the abscissa-axes indicates the atom names of the specified nucleotides.

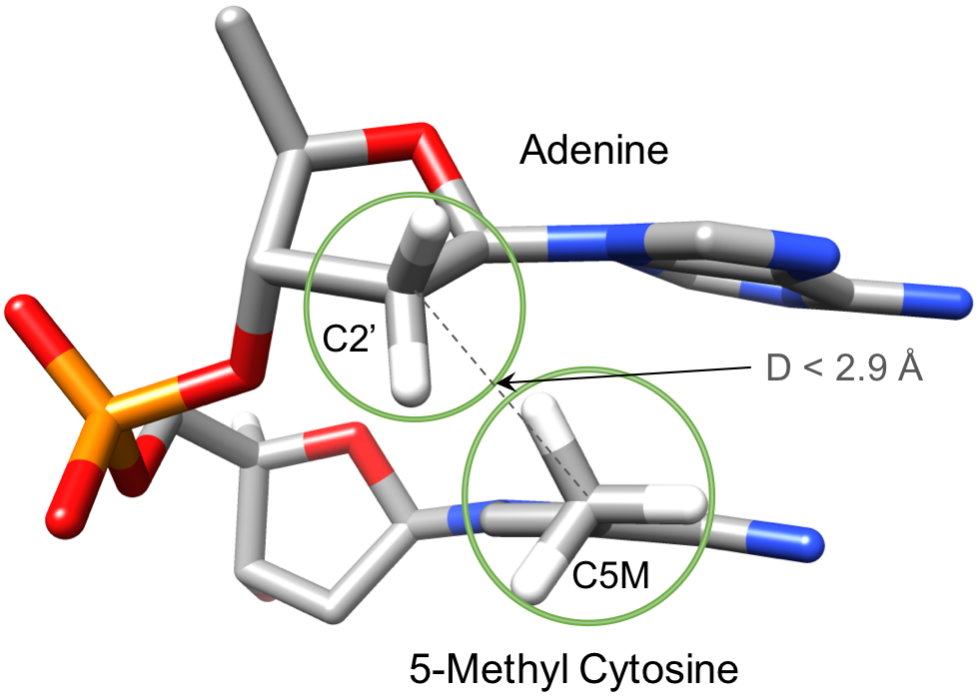

Figure S5. Potentially present steric clash between methyl group of 5-methylcytosine and C2'-atom of sugar phosphate backbone for AMGT-DNA, during molecular underwinding.

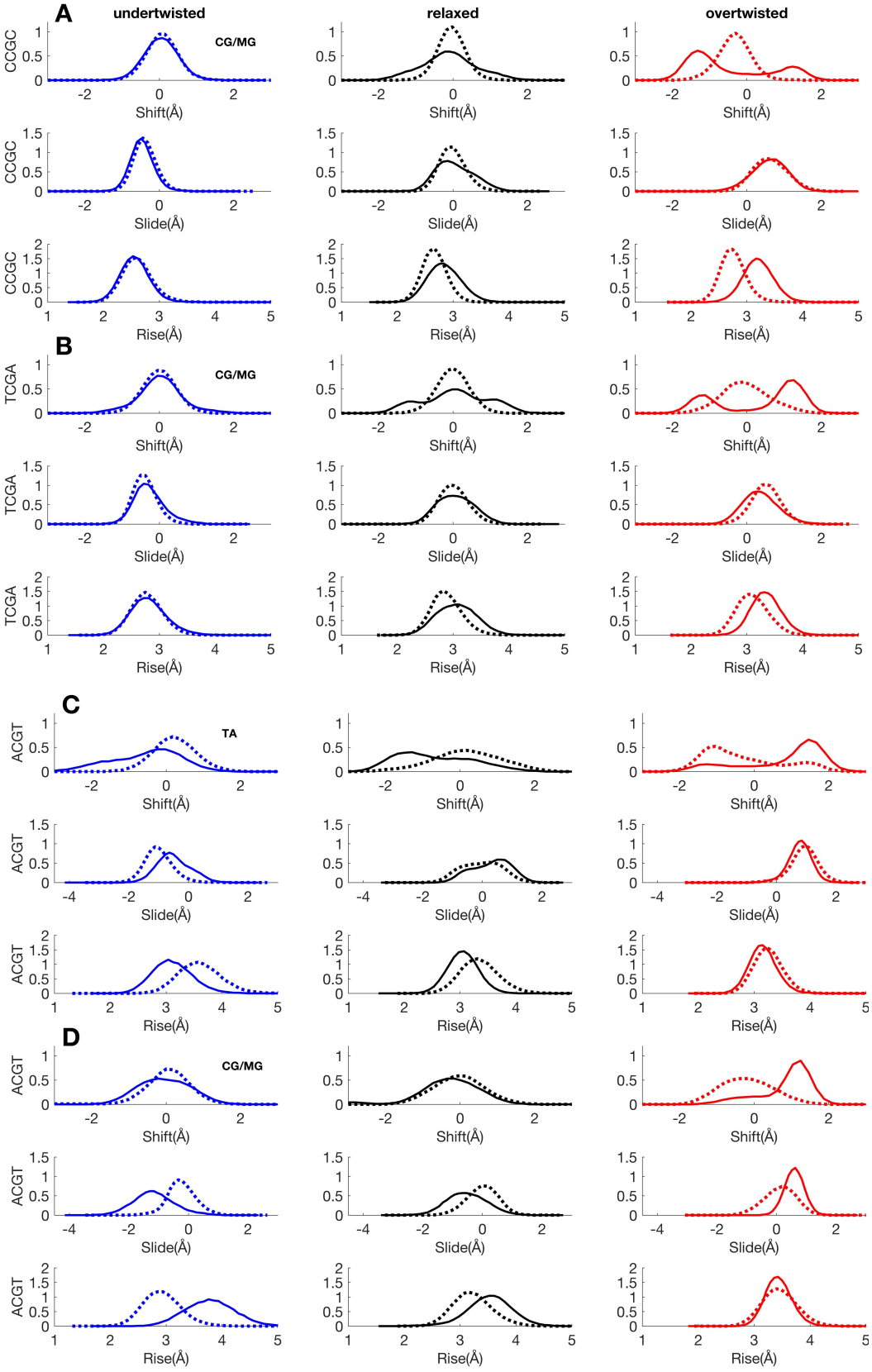

Figure S6. Normalized distributions of translational inter-base pair helical variables for the steps most affected by the imposed twist in each oligomer: CpG/MpG in (A) CCGC/CMGC-DNAs and (B) TCGA/TMGA-DNAs and TpA (C) and CpG/MpG (D) in ACGT/AMGT-DNAs. The dinucleotides from methylated oligomers depicted with dotted lines, and from nonmethylated – with bold lines.

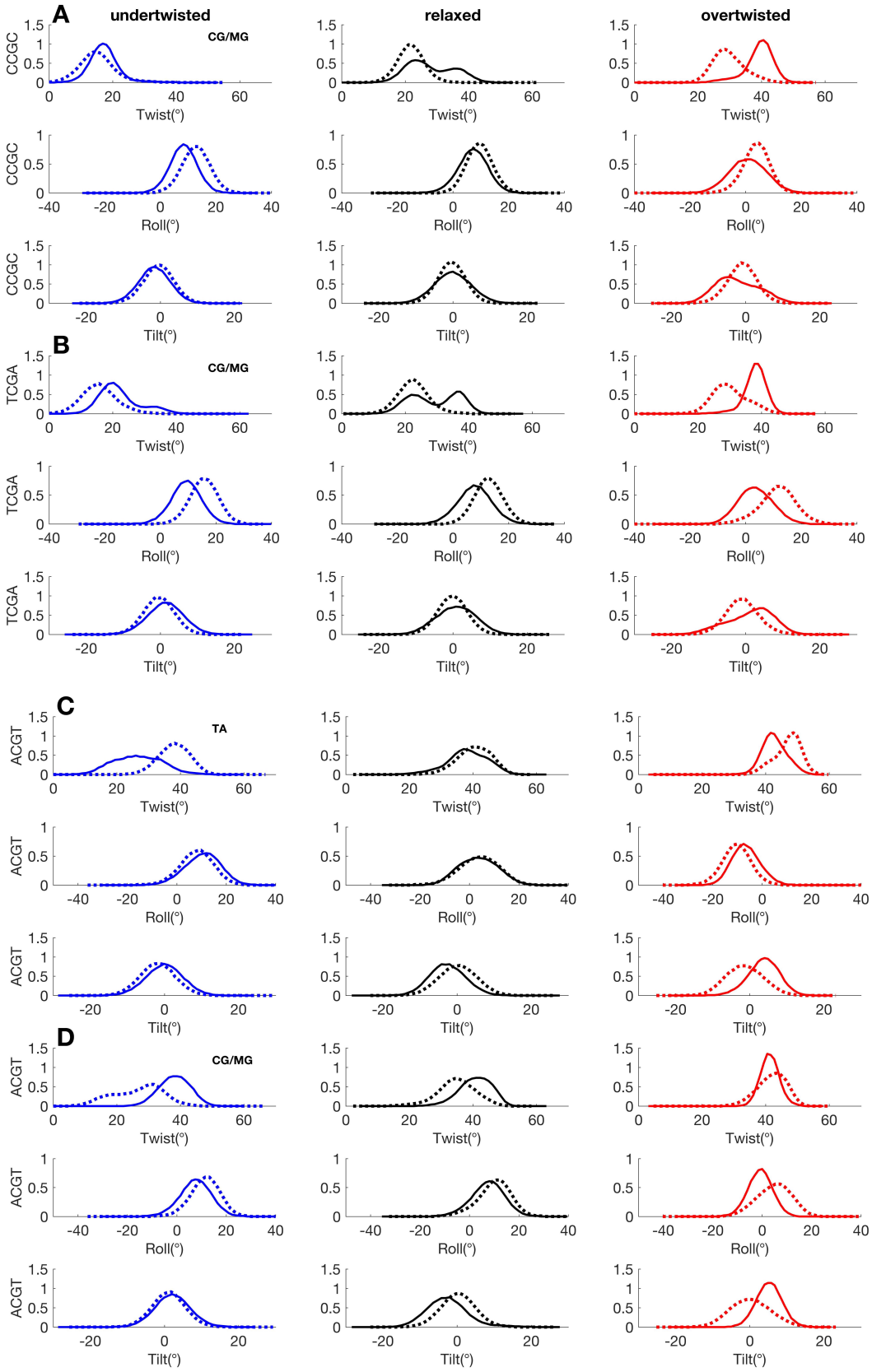

Figure S7. Normalized distributions of rotational inter-base pair helical variables for the steps most affected by the imposed twist in each oligomer: CpG/MpG in (A) CCGC/CMGC-DNAs and (B) TCGA/TMGA-DNAs and (C) TpA and (D) CpG/MpG in ACGT/AMGT-DNAs. The dinucleotides from methylated oligomers depicted with dotted lines, and from nonmethylated – with bold lines.

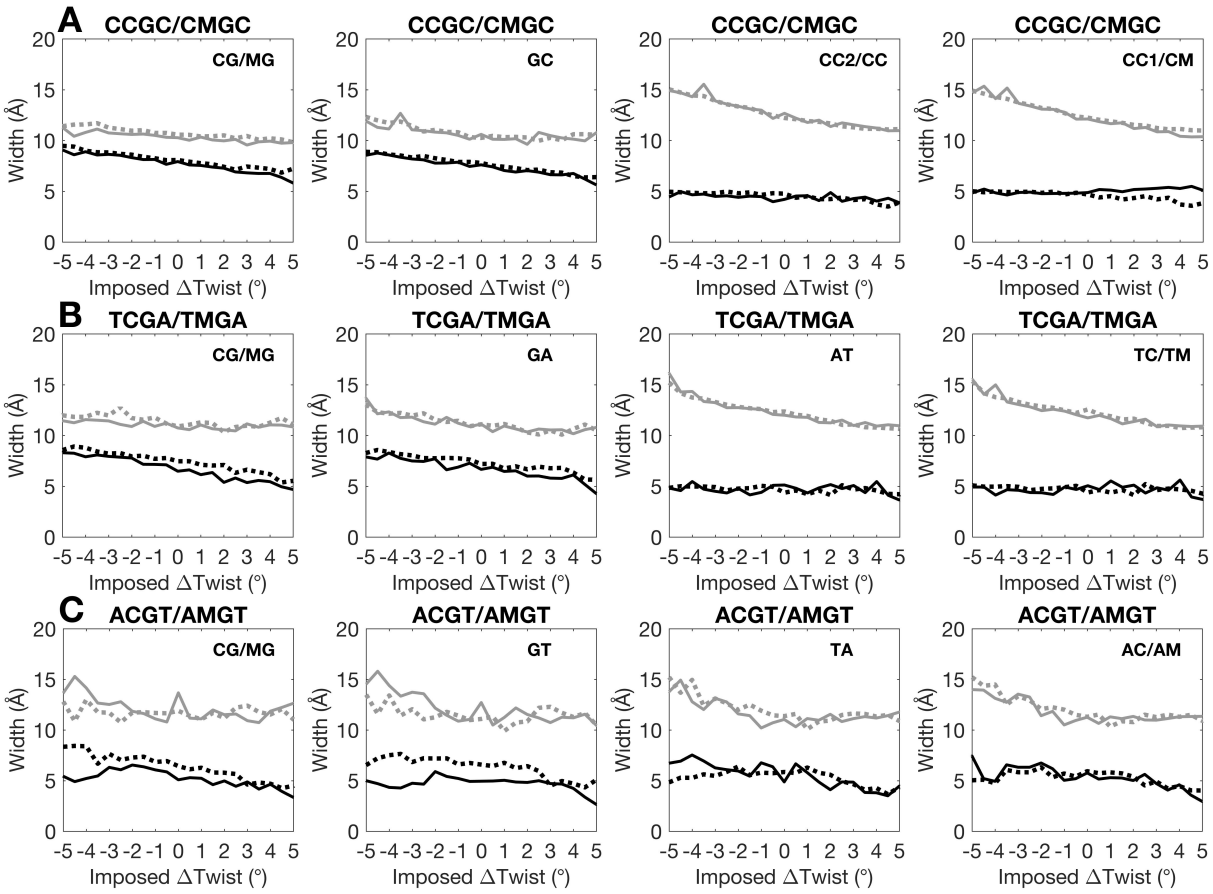

Figure S8. Changes in minor groove width (black) and major groove width (grey) for all the dinucleotide steps for the six oligomers as a function of the imposed twist: the dinucleotides from methylated oligomers depicted with dotted lines, and from nonmethylated – with bold lines.

References

1 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
2 R. Lavery, K. Zakrzewska and H. Sklenar, *Comput. Phys. Commun.*, 1995, **91**, 135–158.
3 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–12.
4 M. James, T. Murtola, R. Schulz, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **2**, 19–25.
5 A. Reymer, K. Zakrzewska and R. Lavery, *Nucleic Acids Res.*, 2018, **46**, 1684–1694.
6 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.
7 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2015, **13**, 55–58.
8 F. Lankaš, T. E. Cheatham, N. Špačáková, P. Hobza, J. Langowski and J. Šponer, *Biophys. J.*, 2002, **82**, 2592–2609.
9 A. T. P. Carvalho, L. Gouveia, C. R. Kanna, S. K. T. S. Wärmländer, J. A. Platts and S. C. L. Kamerlin, *Epigenetics*, 2014, **9**, 1604–1612.
10 P. Mark and L. Nilsson, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.
11 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, 1984, **81**, 3684–3690.
12 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
13 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, 1997, **18**, 1463–1472.
14 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
15 S. C. Harvey, R. K.-Z. Tan and T. E. Cheatham, *J. Comput. Chem.*, 1998, **19**, 726–740.
16 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenbergl, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
17 R. Lavery, J. H. Maddocks, M. Pasi and K. Zakrzewska, *Nucleic Acids Res.*, 2014, **42**, 8138–49.
18 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
19 R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute and K. Zakrzewska, *Nucleic Acids Res.*, 2009, **37**, 5917–5927.