Supporting Information

Water-bridged hydrogen bond formation between 5-hydroxylmethylcytosine (5-hmC) and its 3′-neighbouring bases in A- and B-form DNA Duplexes

Rui Wang, Srivathsan V. Ranganathan, Vibhav A. Valsankar, Stephanie M. Magliocco, Fusheng Shen, Alan Chen and Jia Sheng*

Department of Chemistry and The RNA Institute, University at Albany, State University of New York, 1400 Washington Avenue, 12222, Albany, NY 12203 United States. Email: jsheng@albany.edu

Experimental Section

Synthesis and deprotection of 5hmC modified DNA oligonucleotides

The 5-hydroxylmethylcytidine phosphoramidite was purchased from Glen Research and the 2-SeMe-Uridine phosphoramidite was purchased from SeNA Research. All the other reagents were standard solutions from Chemgenes Corporation. All the phosphoramidites were dissolved in acetonitrile to a concentration of 0.1 M. The DNA oligonucleotides were chemically synthesized at the 1.0-μmol scale by solid phase synthesis. The oligonucleotides were prepared in DMTr-on form. After synthesis, the 5hmC-DNA were cleaved from the solid support and fully deprotected with 30% ammonium hydroxide at 75 °C for 17 hours. The ammonium was removed by Speed-Vac concentration and the solution was analyzed and purified by HPLC.

HPLC purification and analysis

The 5hmC modified DNA oligonucleotides were purified by reverse phase HPLC in both DMTr-on and DMTr-off forms. Purification was carried out using a 21.2 x 250 mm Zorbax, RX-C8 column at a flow rate of 6 ml/min. Buffer A consisted of 20 mM triethylammonium acetate (TEAAc, pH 7.1), while buffer B contained 50% aqueous acetonitrile and 20 mM

TEAAc, pH 7.1. The DMTr-on oligonucleotides were eluted with up to 70% buffer B in 20 min in a linear gradient, while the DMTr-off oligonucleotides were eluted with up to 50% of buffer B in a linear gradient in the same period of time. The collected fractions were lyophilized; the purified compounds were re-dissolved in water. The 5'-DMT groups were removed by the treatment of 3% trichloroacetic acid solution for 5 min, followed by neutralization to pH 7.0 with a freshly made aqueous solution of triethylamine (1.1 M). The trityl-off versions of the DNA oligos were purified again by the same HPLC system. The products were lyophilized and re-dissolved in water with 1mM final concentration.

Crystallization and diffraction data collection

DNA samples (0.5 mM duplex) were heated to 80 °C for 3 minutes, cooled slowly to room temperature, and placed at 4 °C overnight before crystallization. Nucleic Acid Mini Screen Kits (Hampton Research), Natrix (Hampton Research) and Nuc-Pro-HTS (Jena Bioscience) were used to screen crystallization conditions at different temperatures using the hanging drop method. Perfluoropolyether was used as cryoprotectant for the crystal mounting. Data was collected under a liquid nitrogen stream at -174°C. All diffraction data was collected at beam lines ALS 8.2.2 and 8.2.1 at Lawrence Berkeley National Laboratory. A number of crystals were scanned to find one with the highest resolution. Data were collected at a wavelength of 1.0 Å. Crystals were exposed for 1 second per image with a 1 degree oscillation angle. All data were processed using HKL2000 and DENZO/ SCALEPACK[1].

Structure Determination and Refinement

The two DNA structures presented here were solved by molecular replacement with PHASER using PDB structure 1Z7I and 1BNA as the search model respectively, followed by the refinement using Refmac. The usual refinement protocol includes simulated

annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction. The stereo-chemical topology and geometrical restraint parameters of DNA/RNA were applied [2]. The topologies and parameters for 2'-SeMe-U and 5hmC were constructed using Jligand [3]. After several cycles of refinement, a number of highly ordered waters were added. Cross-validation [4] with a 5-10% test set was monitored during the refinement. The σA-weighted maps [5] of the $(2m|Fo| - D|Fc|)$ and the difference $(m|Fo| - D|Fc|)$ density maps were computed and used throughout the model building.

Molecular Dynamics Simulations

To study the effect of the 3' nucleoside on the behavior of hm^{5}C modification in MD simulations, we developed AMBER[6] type force-field parameters for the atoms of the modified nucleoside. We used the online RESP charge-fitting server, R.E.D.S., [7] to obtain the partial charges on the atoms. The geometry of the modified nucleoside was energy minimized, and Hartree-Fock level theory and 6-31G* basis-sets were employed to arrive at a set of partial charges.[8] AMBER99 force-field parameters were used for bonded interactions[6], and AMBER99 parameters with Chen-Garcia corrections were used for LJ interactions.[9]

We studied the modified nucleoside in the context of both A-form and B-form DNA with the 3' nucleoside as Adenine and Guanine, totaling four different simulation systems. The unmodified DNA duplex was constructed in A/B-form using Nucleic Acid Builder (NAB) suite of AMBER 11 package. Using WebMo, we performed mutations such as, C to hm^{5}C and G to A, to get the four different DNA duplexes for MD studies.

Molecular dynamics simulations were performed using Gromacs-4.6.3 package.[10] The simulation system included the DNA duplex in a solution of 1M KCl solution in a 3D periodic box. The box size was 7 x 7 x 7 nm^3 containing 200 K$^+$ ions, 186 Cl$^-$ ions and roughly 10000 water molecules. The system was subjected to energy minimization to prevent any overlap of atoms, followed by a 20 ns production simulation. The MD simulations incorporated leap-frog algorithm with a 2 fs timestep to integrate the equations of motion. The system was maintained at 300K and 1 bar, using the velocity rescaling thermostat[11] and Parrinello-Rahman barostat,[12] respectively. The long-ranged electrostatic interactions were calculated using particle mesh Ewald (PME)[13] algorithm with a real space cut-off of 1.2 nm. LJ interactions were also truncated at 1.2 nm. TIP3P model[14] was used represent the water molecules, and LINCS[15] algorithm was used to constrain the motion of hydrogen atoms bonded to heavy atoms. Co-ordinates of the DNA molecule were stored every 1 ps for further analysis. Analysis of the simulation trajectory included calculation of pair-distances to identify interacting atoms, and dihedral angles to assess the orientation of the hydroxyl group of the modified 5hmC base. Gromacs routines, such as g_dist and g_angle were used for this purpose.
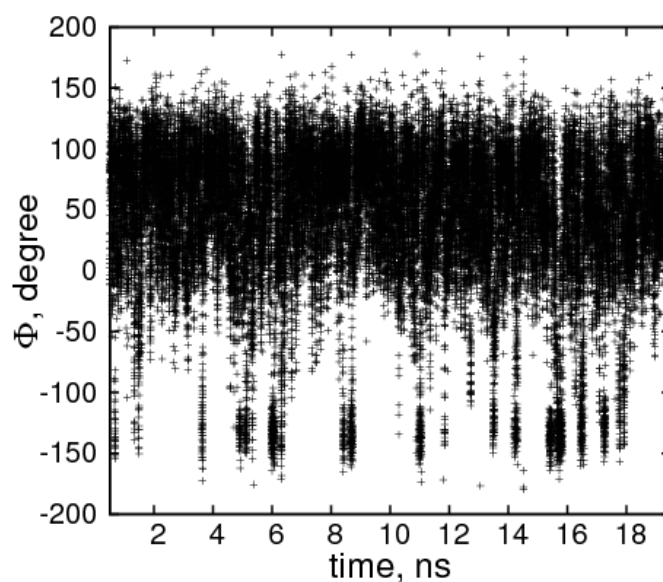
Fig. S1 The time series data for the C6-C5-C7-O7 dihedral angle of the 5hmC base in the B-form DNA with G as the 3' base. The data shows adequate sampling with 10+ transitions between the 3' and 5' side orientation of the hydroxyl group.

References:

1. Z. Otwinowski and W. Minor, *Method Enzymol.*, 1997, **276**, 307.
2. G. Parkinson, J. Vojtechovsky, L. Clowney, A. T. Brunger and H. M. Berman, *Acta. Crystallogr. D Biol. Crystallogr.*, 1996, **52**, 57-64.
3. A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, *Acta. Crystallogr. D Biol. Crystallogr.*, 2012, **68**, 431-440.
4. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, *Acta. Crystallogr. D Biol. Crystallogr.*, 1998, **54**, 905-921.
5. R. J. Read, *Acta. Cryst. A*, 1986, **42**, 140-149.
6. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.
7. F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7821.
8. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *J. Am. Chem. Soc.*, 1993, **115**, 9620-9631.
9. A. A. Chen and A. E. García, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 16820-16825.
10. B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory. Comput.*, 2008, **4**, 435-447.
11. G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
12. H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.

13. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
14. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
15. B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463-1472.