Electronic supplementary information for

A Computational Investigation on the Substrate Preference of Ten-Eleven-Translocation 2 (TET2)

Junyan Lu[a#], Lulu Hu[b,c#], Jingdong Cheng[b#], Dong Fang[d], Chen Wang[a], Kunqian Yu[a], Hualiang Jiang[a], Qiang Cui[d], Yanhui Xu[b,c*] and Cheng Luo[a*]

[a]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

[b]Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai 200032, China.

[c]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China.

[d]Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin-Madison, 1101 University Ave, Madison, WI 53706, USA

Supplementary Methods

Relative binding free energy calculations

Free energy simulations with the Bennett Acceptance Ratio (BAR) method were used to calculate energy changes for mutating 5mC to 5hmC or 5fC in DNA duplex or in catalytic pocket[1, 2]. All the free energy simulations were performed in Gromacs 4.5.5 package[3]. The initial coordinates for the free energy simulations were taken from the

final frame of the canonical MD trajectories. To mutate 5mC to other modified bases in DNA duplex and TET2 catalytic pocket, a following scheme was used. Firstly, the electrostatics on the methyl group of 5mC was linearly turned off from λ=0 to λ=1 using 10 windows ($\Delta\lambda$=0.1). Then the methyl group was alchemically changed to a hydroxyl group or a formyl group by transforming the Lennard-Jones interactions of the exnihilated or annihilated atoms. During the Lennard-Jones transformations, 20 windows ($\Delta\lambda$=0.05) were used to achieve good convergence and soft core potentials were used to avoid endpoint errors[4]. At last, the charges on the hydroxyl or the formyl group was linearly turned on using 10 windows ($\Delta\lambda$=0.1). As the charges on the atoms of the cytosine ring is slightly different in the three modified bases, they were also transformed to the corresponding value in the target base at the same time. For each window in the electrostatic and Lennard-Jones transformations, a 2 ns MD simulation was performed (the first 200 ps for each window was regarded as the equilibration process and did not used in the analysis step). After the simulations, the free energy changes (ΔG_3 and ΔG_4) as well as the statistical errors were calculated using the *g_bar* program in Gromacs 4.5.5 package. To validate the free energy calculations, we also performed reversed transformations (from 5hmC to 5mC ,from 5fC to 5mC, and from 5fC to 5hmC) using the same scheme describe above.

Supplementary Figures

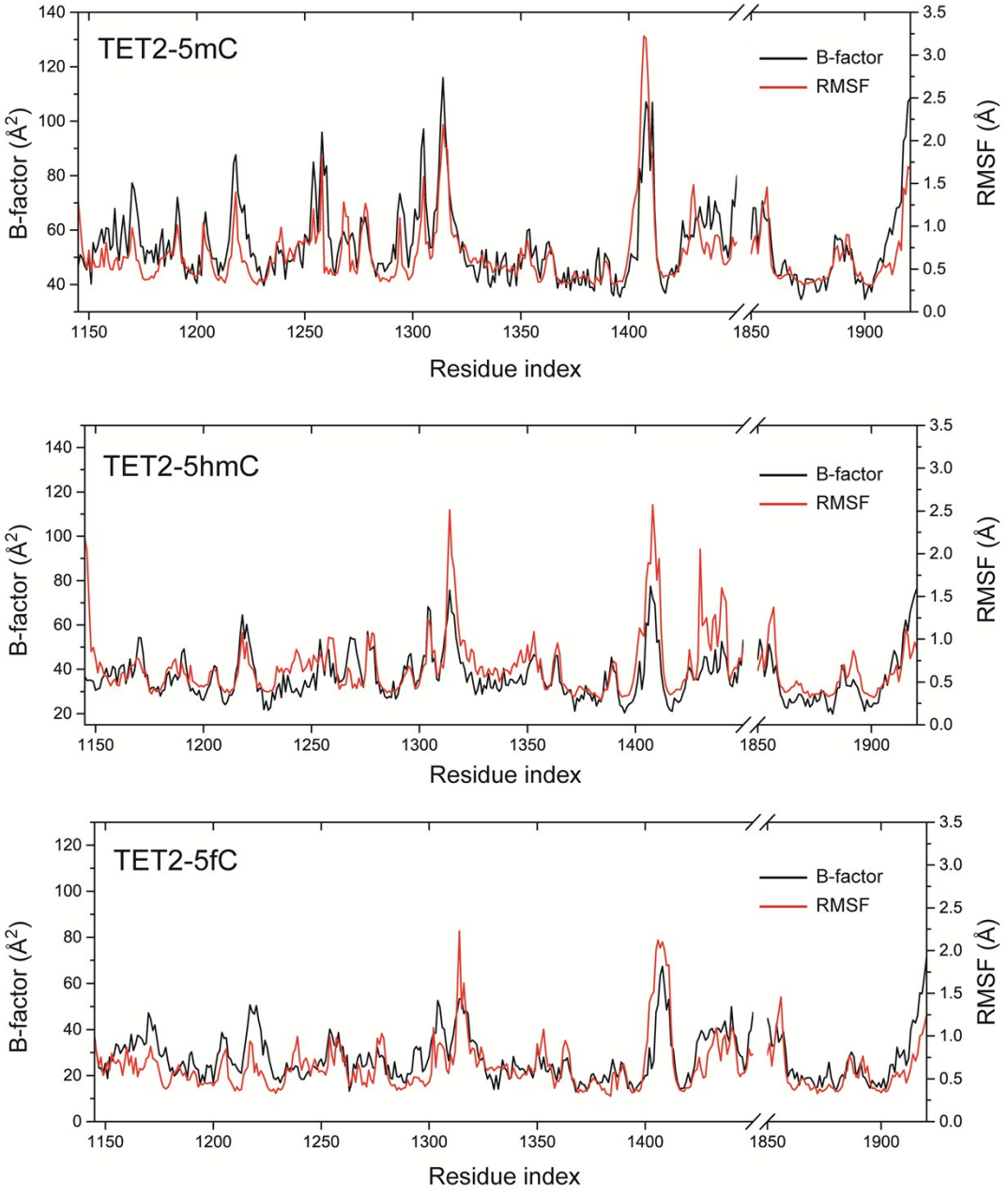
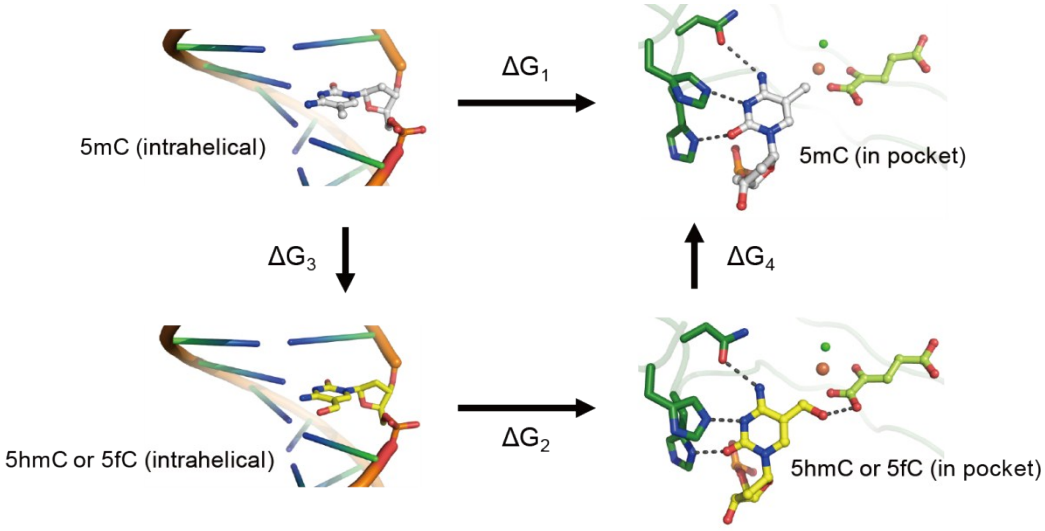


Figure S1. Comparison of RMSF (root mean square fluctuation) values with crystallographic B-factor values. The RMSF values of Cα of protein were calculated from the MD trajectory of each TET2-substrate complex. B-factor values for the unresolved residues in the crystal structures are absent.



$$\Delta\Delta G_{(\,5mC\,to\,5hmC\,)} = \Delta G_1 - \Delta G_2 = \Delta G_3 + \Delta G_4 = 0.39\ kcal/mol$$

$$\Delta\Delta G_{(\,5mC\,to\,5fC\,)} = \Delta G_1 - \Delta G_2 = \Delta G_3 + \Delta G_4 = 1.85\ kcal/mol$$

Figure S2. Thermodynamic cycle for the calculation of free energy differences for the processes that TET2 binds to various DNA substrates

A thermodynamic cycle was designed to calculated the relative free energy differences for TET2 to bind and flip a 5mC, 5hmC or 5fC from DNA double helix into its catalytic pocket, denoted as $\Delta\Delta G = \Delta G_1 - \Delta G_2$. As $\Delta\Delta G$ is a state function, it can be derived by a hypothetical paths for mutating an intrahelical 5mC to an intrahelical 5hmC or 5fC (ΔG_3) and a 5mC in the pocket to a 5hmC or 5fC (ΔG_4). The free energy difference can then be calculated by $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_3 + \Delta G_4$.

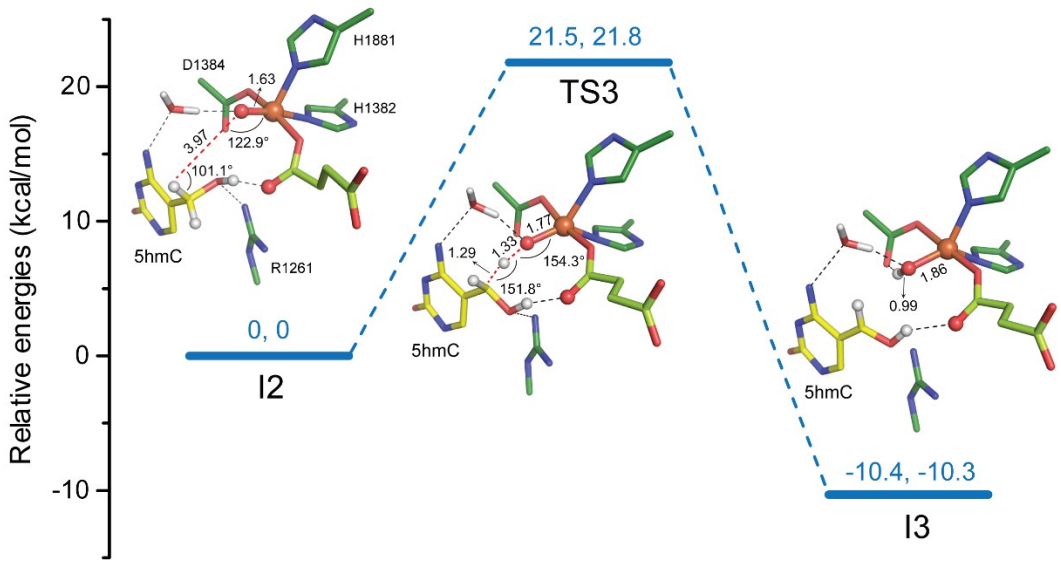


Figure S3. Energy profile and critical structures for hydrogen abstraction from 5hmC by the Oxo group lies trans to H1382. Relative energies (given as ΔE, ΔG) for the Fe(IV)-oxo intermediates (I2), transition states (TS) and Fe(III)-OH intermediates (I3) and their structures are shown.

Supplementary Tables

Table S1. Relative binding free energies of 5mC to 5hmC and 5fC from free energy perturbation (FEP) calculations

FEP direction	ΔG in protein (kcal/mol)	ΔG in DNA helix (kcal/mol)	ΔΔG binding (kcal/mol)
5mC to 5hmC	-47.69 ± 0.07	-48.08 ± 0.06	0.39 ± 0.07
5hmC to 5mC	47.92 ± 0.05	48.52 ± 0.06	-0.60 ± 0.06
5mC to 5fC	-30.75 ± 0.06	-32.60 ± 0.15	1.85 ± 0.15
5fC to 5mC	31.11 ± 0.14	32.60 ± 0.13	-1.49 ± 0.13
5hmC to 5fC	15.75 ± 0.06	14.93 ± 0.11	0.82 ± 0.11
5fC to 5hmC	-15.74 ± 0.21	-15.23 ± 0.18	-0.51 ± 0.21

ΔΔG binding is calculated as ΔG in protein minus ΔG in DNA helix. The absolute values of ΔΔG binding calculated from forward transformations (5mC to 5hmC, 5mC to 5fC and 5hmC to 5fC) are close to that calculated from reversed transformations (5hmC to 5mC, 5fC to 5mC and 5fC to 5hmC), which indicates the FEP calculations are reliable.

Table S2. Spin densities of the catalytic center atoms in different substrates and reaction states.

		C5A	H1	O	FE
Reactant	5mC	0.00	0.00	-0.21	4.24
	5hmC	0.00	0.00	-0.18	4.27
	5fC	0.00	0.00	-0.20	4.25
TS1	5mC	0.00	0.00	0.16	4.23
	5hmC	0.00	0.00	0.13	4.27
	5fC	0.00	0.00	0.20	4.21
I1	5mC	0.00	0.00	0.03	3.81
	5hmC	0.00	0.00	0.04	3.81
	5fC	0.00	0.00	0.04	3.81
TS2	5mC	0.00	0.00	0.01	3.90
	5hmC	0.00	0.00	0.02	3.93
	5fC	0.00	0.00	0.02	3.91
I2	5mC	0.00	0.00	0.58	3.23
	5hmC	0.00	0.00	0.55	3.26
	5fC	0.00	0.00	0.61	3.20
TS3	5mC	-0.34	0.03	0.08	4.04
	5hmC	-0.22	0.02	0.02	4.01
	5fC	-0.23	0.03	0.09	4.00
I3	5mC	-0.76	0.00	0.25	4.29
	5hmC	-0.54	0.00	0.31	4.25
	5fC	-0.67	0.00	0.29	4.29
TS4	5mC	-0.34	0.00	0.13	4.05
	5hmC	-0.58	0.00	0.15	4.28
	5fC	-0.35	0.00	0.13	4.10
Product	5mC	0.00	0.00	0.00	3.82
	5hmC	0.00	0.00	0.00	3.82
	5fC	0.00	0.00	0.00	3.83

*C5A is the carbon atom of 5-substitution groups on different bases; H1 is the abstractable hydrogen on 5-subsitution groups; O is the oxygen atom that directly binds to the iron; Fe is the iron atom in the reaction center.

References

1. C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245-268.
2. S. E. Boyce, D. L. Mobley, G. J. Rocklin, A. P. Graves, K. A. Dill and B. K. Shoichet, *J. Mol. Biol.*, 2009, **394**, 747-763.
3. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson and D. van der Spoel, *Bioinformatics*, 2013, **29**, 845-854.
4. V. Gapsys, D. Seeliger and B. L. de Groot, *J. Chem. Theory Comput.*, 2012, **8**, 2373-2382.