Supplemental Information

Structure of *Naegleria* Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA

Hideharu Hashimoto[1,$], June E. Pais[2,$], Nan Dai[2], Ivan R. Corrêa Jr[2], Xing Zhang[1], Yu Zheng[3], Xiaodong Cheng[1,*]

[1] Departments of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

[2] New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

[3] RGENE Inc., 953 Indiana Street, San Francisco, CA, 94107

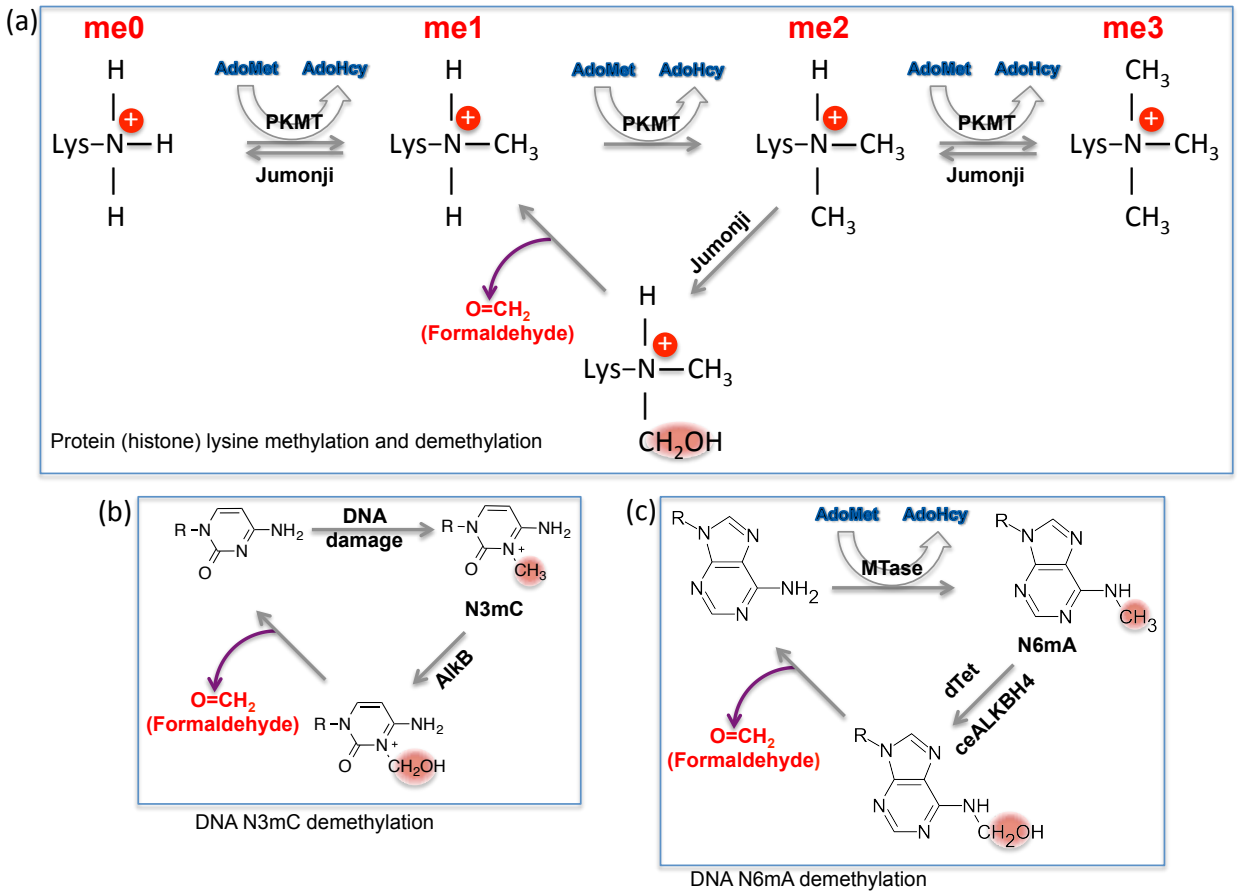
[$] Co-first author

Correspondence and requests for materials should be addressed to:

Xiaodong Cheng (Email: xcheng@emory.edu; Tel: +1 404 727 8491; Fax: +1 404 727 3746)

Email addresses for all authors:

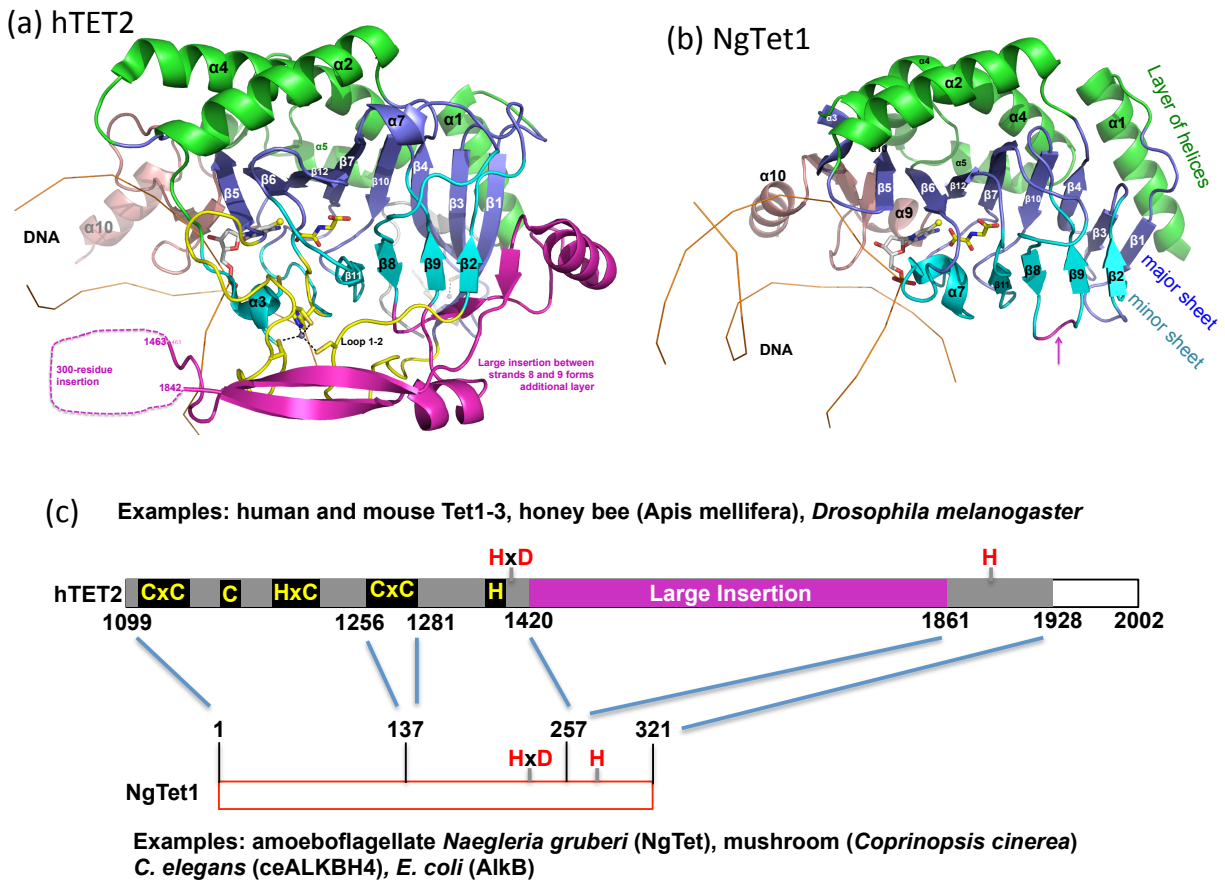
HH (hhashi3@emory.edu), JEP (pais@neb.com), ND (Dain@neb.com), IRC (correa@neb.com), XZ (xzhan02@emory.edu), YZ (yu.zhengyu@gmail.com), XC (xcheng@emory.edu)



Supplemental Figure S1. General mechanism of Fe(II)- and α-ketoglutarate-dependent dioxygenases including demethylases of amino (N)-methylation.

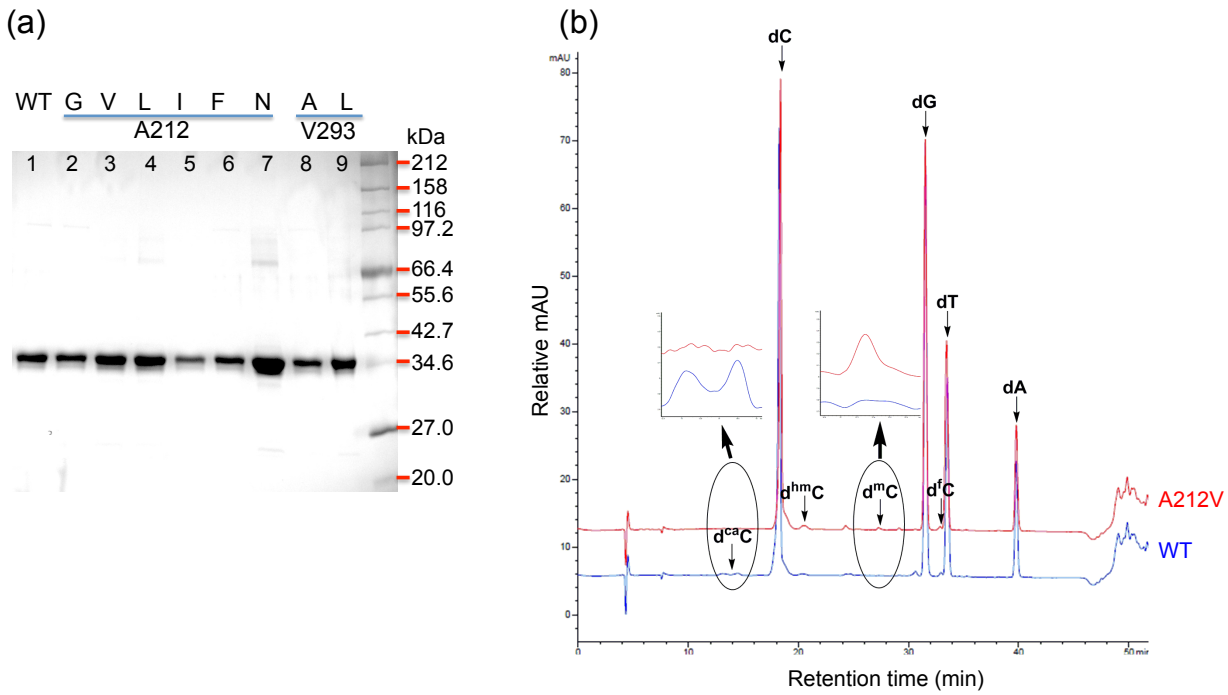
(a) Overview of protein lysine methylation by AdoMet-dependent methyltransferases and demethylation reactions catalyzed by Jumonji dioxygenases. For protein lysine demethylation, the Fe(II)- and α-ketoglutarate-dependent Jumonji dioxygenases (1) generate a hydroxymethyl intermediate (N-CH_2OH) for each reaction that subsequently decomposes to release a formaldehyde spontaneously (without additional enzymatic activities) and the demethylated lysine (with one methyl group removed). In fact, in search of enzymes capable of reversing methylated lysines in histones, the purification of JHDM1, the first identified Jumonji domain-containing histone demethylase, used a biochemical assay based on the detection of formaldehyde, one of the predicted reaction products (2). (**b**) Demethylation of N3-

methylcytosine (N3mC) by AlkB dioxygenase is involved in the direct reversal of alkylation damage, and results in the production of formaldehyde and unmodified cytosine. The same basic mechanism applies to demethylation of other N-methylated nucleic acids (e.g. N3-methylthymine, N1-methyladenine and N6-methyladenine) (3-6). (**c**) Model of DNA N6-adenine methylation by methyltransferase (MTase), generating N6-methyladenine (N6mA), and demethylation reaction product recently discovered in worms, flies and green algae (7-9). The same basic mechanism applies to demethylation of N6mA in mRNA by ALKBH5 and FTO (10,11).

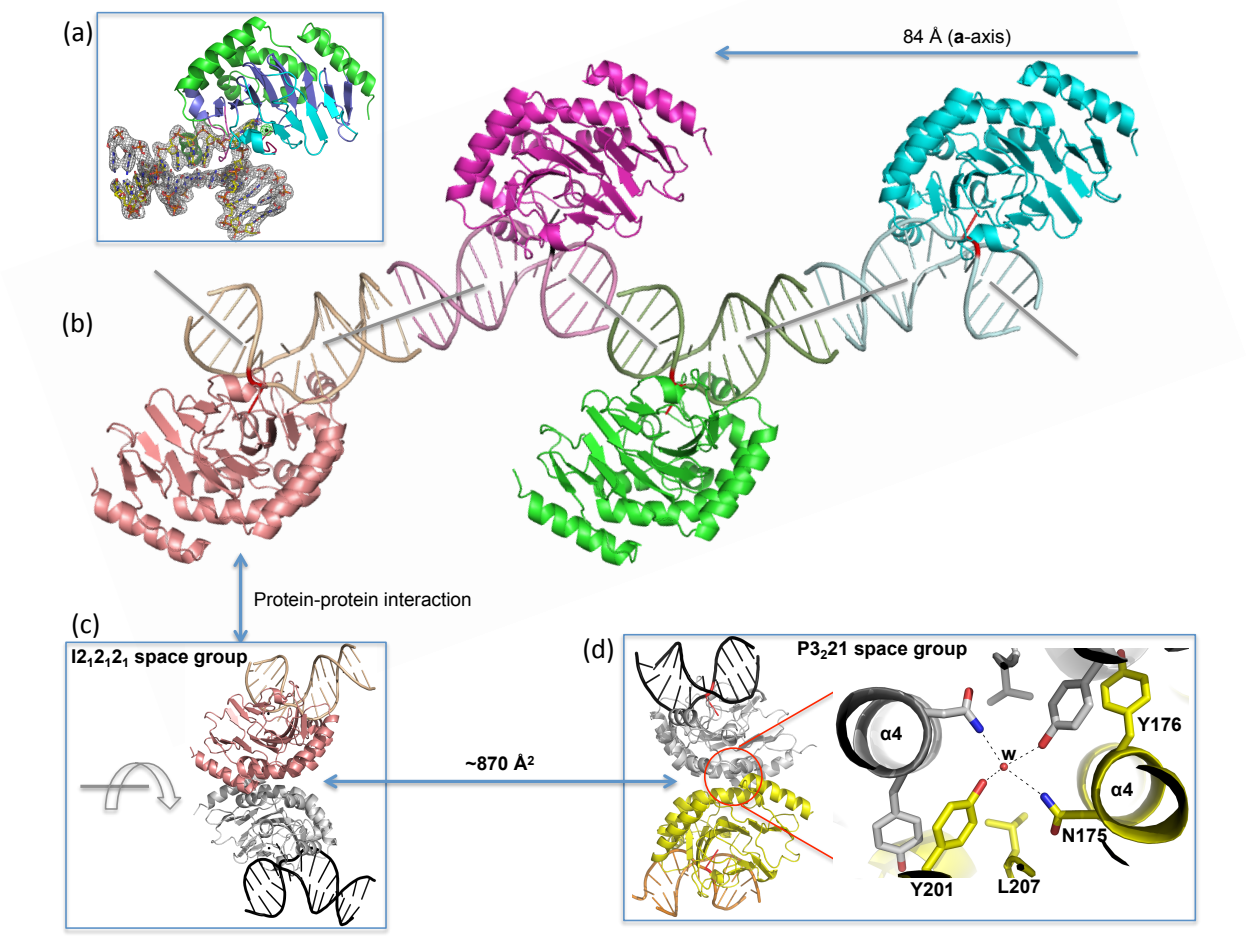


Supplemental Figure S2. Structures of the Tet enzymes.

(**a**) Structure of human TET2-5mC DNA complex (PDB 4NM6). The secondary structure elements are labeled according to the NgTet1 structure (panel b). Note the large insertion in hTET2 between strands 8 and 9 (magenta), which is indicated by a magenta arrow in panel b. (**b**) Structure of NgTet1-5mC DNA complex (PDB 4LT5). The NgTet1 protein folds in a three-layered jelly-roll structure. (**c**) Schematic representation of human TET2 (hTET2) C-terminal catalytic domain and NgTet1. The 5mC dioxygenases can be divided into two sub-families based on size: the larger, multi-domain Tet proteins including mammalian Tet1-3 (12), and Tet homologs from honey bee (*Apis mellifera*) (13) and *Drosophila melanogaster* (CG43444) (14), and the smaller, single domain proteins from *Naegleria* (NgTet1) and mushroom (*Coprinopsis cinerea*) (15). The *Drosophila* Tet homolog catalyzes demethylation of N6mA in DNA both *in vivo* and *in vitro* as well as 5mC *in vitro* (8). In addition, single domain proteins of the AlkB family including *E. coli* AlkB and *C. elegans* ALKBH4 (F09F7.7) are active on DNA N6mA demethylation (5-7).



Supplemental Figure S3. (**a**) A Novex® 10–20% Tris-Glycine polyacrylamide gel (ThermoFisher Scientific; EC61352) showing the proteins used for activity. NEB protein ladder (New England Biolabs; P7702) was used as molecular weight markers. (**b**) LC–MS traces of a sample reaction mix on the 5mC–containing DNA with A212V (red) or wild-type (WT) enzyme (blue). Arrows indicate peaks of 5mC, 5hmC, 5fC and 5caC as well as A, T, G, C. Identities of the peaks are confirmed by comparing the retention time with the standard as well as by mass spectrometry.



Supplemental Figure S4. NgTet1 protein-protein interactions in crystals.

(**a**) In both crystal forms ($I2_12_12_1$, and $P3_221$), the DNA is bound to the inner β-sheet surface of the protein with substantial protein-induced distortions from B-form DNA. The phosphate backbone flanking the modified cytosine is kinked ~60° and concurrently, the modified nucleotide flips out. Electron density 2Fo – Fc, contoured at 1σ above the mean, is shown for the entire 14-bp DNA with a flipped out 5hmC in the $I2_12_12_1$ space group. (**b**) In the $I2_12_12_1$ space group, the bent duplex stacks head-to-head with one neighboring DNA molecule at each end forming a zigzag superhelix. The axis of the superhelix formed by the assembly of the oligonucleotide is parallel to the crystal **a**-axis. The assembled superhelices are connected via NgTet1 molecules through the protein-to-protein interfaces along the **b** and **c** axes that complete

the formation of the crystal. (**c**) One of the protein interfaces in the $I2_12_12_1$ space group is mediated by the outer helical surface. (**d**) In the $P3_221$ space group formed by NgTet1-5mC complex with a 12-bp DNA plus a 5'- thymine overhang, the crystallographic asymmetric unit contains two NgTet1-DNA complexes, using the same outer helical interface as in panel c. The interface of ~869 $Å^2$ is within the range (584-2786 $Å^2$), though in the lower end, of the homodimer interfaces examined (16). The interface involves two pairs of aromatic tyrosines (Y176 and Y201), hydrophobic L207 and I190 (not shown), and a network of hydrogen bonds formed by the side chain amine of N175 and hydroxyl oxygen atom of Y201. Although the functional significance, if any, of the NgTet1 'crystallographic dimer' is unclear, the NgTet1-DNA packing is very similar to that of the HinP1I endonuclease-DNA interaction (17) and DNA bridge factors that induce DNA condensation by two DNA binding motifs located on the opposite surfaces of a bridge dimer (18). There are also known DNA methyltransferases exist as a dimer: mammalian Dnmt3a-3L complex (19), group β DNA amino (N4-cytosine or N6-adenine) methyltransferases [PvuII (20), RsrI (21), and MboIIA (22)], but no nucleic acid dioxygenases have been found to act as a dimer.

REFERENCES

1. Shi, Y.G. and Tsukada, Y. (2013) The discovery of histone demethylases. *Cold Spring Harb. Perspect. Biol.*, **5**(9), pii: a017947. doi: 10.1101/cshperspect.a017947.
2. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P. and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, **439**, 811-816.
3. Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. and Sedgwick, B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature*, **419**, 174-178.
4. Falnes, P.O., Johansen, R.F. and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature*, **419**, 178-182.
5. Li, D., Delaney, J.C., Page, C.M., Yang, X., Chen, A.S., Wong, C., Drennan, C.L. and Essigmann, J.M. (2012) Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the Escherichia coli adaptive response protein AlkB. *J. Am. Chem. Soc.*, **134**, 8896-8901.
6. Zhu, C. and Yi, C. (2014) Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. *Angewandte Chemie*, **53**, 3659-3662.
7. Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., Hsu, C.H., Aravind, L., He, C. and Shi, Y. (2015) DNA Methylation on N(6)-Adenine in C. elegans. *Cell*, **161**, 868-878.
8. Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., Li, C., Liu, B., Luo, Y., Zhu, Y., Zhang, N., He, S., He, C., Wang, H. and Chen, D. (2015) N(6)-methyladenine DNA modification in Drosophila. *Cell*, **161**, 893-906.
9. Fu, Y., Luo, G.Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C., Weng, X., Ji, Q., Mets, L. and He, C. (2015) N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell*, **161**, 879-892.
10. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G. and He, C. (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.*, **7**, 885-887.
11. Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H., Lu, Z., Bosmans, R.P., Dai, Q., Hao, Y.J., Yang, X., Zhao, W.M., Tong, W.M., Wang, X.J., Bogdan, F., Furu, K., Fu, Y., Jia, G., Zhao, X., Liu, J., Krokan, H.E., Klungland, A., Yang, Y.G. and He, C. (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell*, **49**, 18-29.
12. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.
13. Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M. and Maleszka, R. (2014) Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol.*, **4**(8). pii: 140110. doi: 10.1098/rsob.140110.

14. Dunwell, T.L., McGuffin, L.J., Dunwell, J.M. and Pfeifer, G.P. (2013) The mysterious presence of a 5-methylcytosine oxidase in the Drosophila genome: possible explanations. *Cell Cycle*, **12**, 3357-3365.
15. Zhang, L., Chen, W., Iyer, L.M., Hu, J., Wang, G., Fu, Y., Yu, M., Dai, Q., Aravind, L. and He, C. (2014) A TET homologue protein from Coprinopsis cinerea (CcTET) that biochemically converts 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.*, **136,** 4801-4804.
16. Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U S A*, 93, 13-20.
17. Horton, J.R., Zhang, X., Maunus, R., Yang, Z., Wilson, G.G., Roberts, R.J. and Cheng, X. (2006) DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.*, **34**, 939-948.
18. Bradley, C.M., Ronning, D.R., Ghirlando, R., Craigie, R. and Dyda, F. (2005) Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat. Struct. Mol. Biol.*, **12**, 935-936.
19. Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, **449**, 248-251.
20. Gong, W., O'Gara, M., Blumenthal, R.M. and Cheng, X. (1997) Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.*, **25**, 2702-2715.
21. Thomas, C.B., Scavetta, R.D., Gumport, R.I. and Churchill, M.E. (2003) Structures of liganded and unliganded RsrI N6-adenine DNA methyltransferase: a distinct orientation for active cofactor binding. *J. Biol. Chem.*, **278**, 26094-26101.
22. Osipiuk, J., Walsh, M.A. and Joachimiak, A. (2003) Crystal structure of MboIIA methyltransferase. *Nucleic Acids Res.*, **31**, 5440-5448.