Nucleoside-driven specificity of DNA Methyltransferase

Madhuri Gade[1], Jasmine M. Gardner[2], Prashant Jain[1], Paola Laurino*[1]

[1]Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa, Japan 904-0495
[2] Merck KGaA, Darmstadt, Germany

Email*: paola.laurino@oist.jp

Abstract

We have studied adenosine binding specificities of two bacterial DNA methyltransferase Taq methyltransferase (M.TaqI), and HhaI methyltransferase (M.HhaI). While these DNA methyltransferases have similar cofactor binding pocket interactions, experimental data showed different specificity for novel cofactors ((SNM) (S-guanosyl-L-methionine (SGM), S-cytidyl-L-methionine (SCM), S-uridyl-L-methionine (SUM)). Protein dynamics corroborate the experimental data on the cofactor specificities. For M.TaqI the specificity for S-adenosyl-L-methionine (SAM) is governed by the tight binding on the nucleoside part of the cofactor, while for M.HhaI the degree of freedom of the nucleoside chain allows the acceptance of other bases. The experimental data proves a catalytically productive methylation by M.HhaI binding pocket for all the SNM. Our results suggest a new route for successful design of unnatural SNM analogues for methyltransferases as a tool for cofactor engineering.

Introduction

Nucleotides and their derivatives are the most common chemical moiety present in natural ligands[1]. Adenosine moiety is the most abundant among the nucleotides and plays a key role in all the living organisms[2, 3]. Adenosine's role as a cofactor or ligand in various cellular processes is fulfilled by promoting specific interactions with proteins and enzymes. These specific interactions allow proteins to distinguish between adenosine and other ligands or nucleotides[2, 4, 5]. Ultimately adenosine specific interactions are exploited by the cell for the fine regulation of enzymes in the context of metabolic processes[4]. Adenosine moiety is overrepresented in cofactors, namely adenosine 5′-triphosphate (ATP), nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD), S-adenosyl methionine (SAM), and coenzyme A (CoA). Although SAM is involved in various biological processes in cells[6-9], its main role is methylation of biomolecules[10]. This methylation is facilitated by a class of enzymes called methyltransferases[11].

Methyltransferase is a large family of enzymes that catalyzes the transfer of a methyl group from SAM to biomolecules (DNA, RNA, protein, and small molecules)[12, 13] (Figure 1a). Methylation of these biomolecules plays a critical role in various processes like gene expression and regulation, signal transduction, gene silencing, and chromatin regulation[14, 15]. One important class of methyltransferases is DNA methyltransferase[16, 17]. For instance, in prokaryotes, DNA methyltransferase has a crucial role in distinguishing between endogenous and exogenous DNA[18]. In mammals, DNA methylation plays a key role in embryonic development and gene regulation[19]. Considering the significance of DNA methylation, it is essential to find an easy method to detect methylation patterns in cells. One approach includes the use of SAM analogues that carry an activated transferring group which will allow the labeling of the substrates in cells and ultimately the study of the methyltransferase's function[20-24]. However, use of SAM analogs remains limited because of the lack of biorthogonality[25, 26] for instance the specific recognition of synthetic cofactor that ultimately allows fine regulation of an engineered methyltransferase.

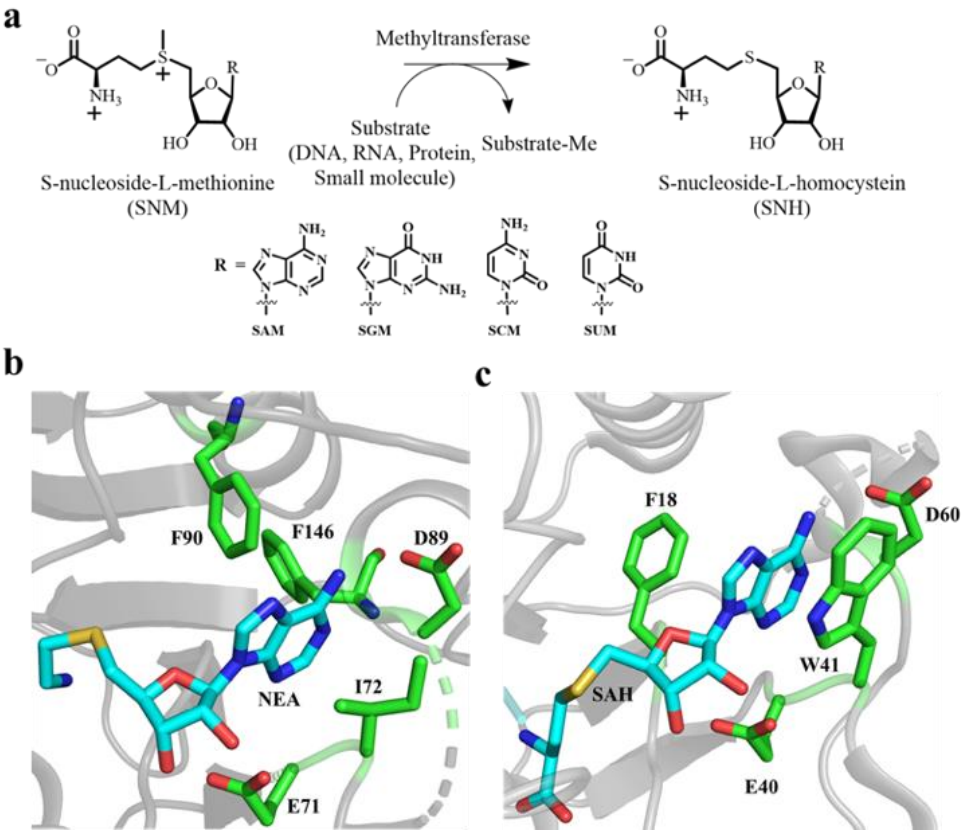

Figure 1. Natural methylation reaction and binding pockets of M.TaqI and M.HhaI DNA methyltransferase. (a) Schematics of transfer of methyl group from cofactor to biomolecules (DNA, RNA, protein, and small molecules) and structures of S-nucleoside methionine (SNM) cofactor used in this study. (b) Binding site residues of M.TaqI interacting with adenosine of 5'-Deoxy-5'-[2-(amino)ethylthio]adenosine (NEA). NEA is in cyan and interacting residues are in green color (PDB ID 1g38). (C) Binding site residues of M.HhaI interacting with adenosine of SAH. SAH is in cyan and interacting residues are in green (PDB ID 2hr1).

As mentioned above, adenosine shows specific interactions with protein. In DNA methyltransferases, adenosine moiety has several sites for hydrogen bonding (acceptors N1, N3, and donor N6 of adenine and ribose hydroxyl group) and pi-pi stacking interaction with adenine aromatic ring[2]. N6 of adenine is primarily involved in hydrogen bonding with the side chain of aspartate residue[3]. It is well documented that ribose hydroxyl forms bidentate interaction with aspartate or glutamate[27]. Indeed, interactions between SAM and DNA methyltransferases at the active site are essential for the specificity of the enzyme.

By taking advantage of these specific interactions, one could think of replacing the adenosine moiety, achieve specificity and ultimately orthogonality for a SAM analogue carrying a different nucleobase. A recent report functionalized adenosine with different lateral group, however no changes of the aromatic core is reported so far[20].

At the same time a detailed understanding of interactions is required to analyse factors governing cofactor discrimination. Recognition between cognate and non-cognate SAM analogs by the DNA methyltransferases is usually elucidated by sequence specificity that allows specific protein-cofactor interactions and steric hindrance in the cofactor binding pocket[28]. Still, it is less understood if enzyme dynamics play a role in differentiating between cofactor analogs.

Herein, we have investigated the adenosine binding site for two DNA methyltransferases by utilizing novel S-nucleoside methionine (SNM) (S-guanosyl-L-methionine (SGM), S-cytidyl-L-methionine (SCM), S-uridyl-L-methionine (SUM)) cofactors[29]. These SNM cofactors carry different nucleotide bases, making them ideal candidates to study the adenosine binding pocket of DNA methyltransferase. This study opens a new avenue for the design of SAM derivatives carrying different nucleobases and their potential to achieve biorthogonality because of substantial changes in the cofactor structure.

Results and Discussion

We tested two representative classes of DNA methyltransferase Taq methyltransferase (M.TaqI), and HhaI methyltransferase (M.HhaI). M.TaqI recognize the DNA sequence TCG**A** methylating N6 of the adenine base[30], while M.HhaI recognizes G**C**GC sequence methylating the C5 cytosine of the cytosine base [31]. In M.TaqI, adenosine N1 and N6 interacts via hydrogen bonding with F90, D89 respectively. Adenine ring forms Van der Waals interaction with the side chain of I72, F90, and F146. The ribose hydroxyl groups have a bidentate interaction with E71[32] (Figure 1b).In the case of M.HhaI adenosine, N6 forms hydrogen bonding with D60. Adenine ring of SAH has pi-pi staking interaction with W41 (parallel) and F18 (perpendicular) (Figure 1b). The ribose hydroxyl groups have bidentate interaction with E40[33]. M.TaqI and M.HhaI have identical key interactions with the adenosine group of SAM, this makes them optimal candidates to investigate the contribution of the enzyme dynamics for cognate and non-cognate cofactor preference.

To investigate the specificity of M.TaqI and M.HhaI for different SNM (SAM, SGM, SCM and SUM) analogs, we utilized the enzymatically synthesized SNM cofactors by hMAT2A[29] .To test the methylation activity of M.TaqI and M.HhaI for the different SNM, we used DNA protection against the restriction enzyme assay[34] (Figure 2a). Modification of DNA with methyl group by methyltransferase prevents fragmentation by corresponding restriction enzymes that are methylation-sensitive[35]. First, we set up the reaction of hMAT2A with NTPs (ATP, GTP, CTP, and UTP) and methionine for one hour then, hMAT2A was inactivated, followed by the addition of DNA substrate and DNA methyltransferase. We used dsDNA substrate for M.TaqI and M.HhaI, with one recognition site for methylation. Using dsDNA substrate and various cofactor (SAM, SGM, SCM, and SUM), we carried out a methylation reaction using M.TaqI at 65 °C and M.HhaI at 37 °C for one hour followed by digestion with restriction enzymes TaqI and HhaI respectively for one hour and run on the agarose gel. This assay demonstrates that M.TaqI is specific for SAM, though 50% activity was observed for SCM too (Figure 2b).

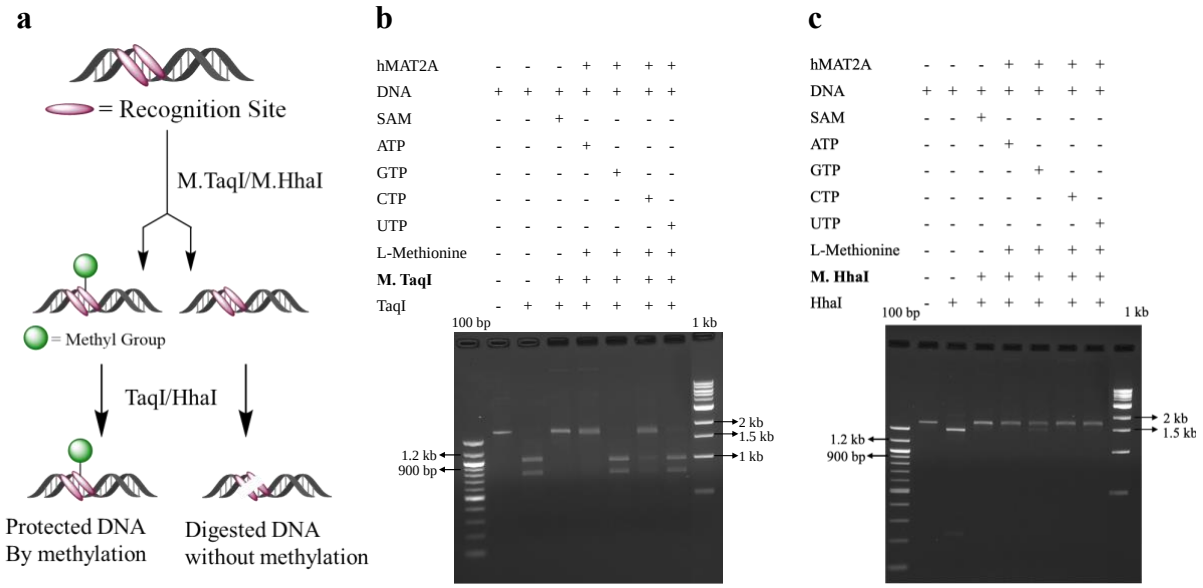

Figure 2. Agarose gel analysis of DNA modification. Enzymatic modification of (a) M.TaqI and (b) M.HhaI byNM cofactors. Agarose gel was 1.5 % run for 1 hr. at 100 V, stained with ethidium bromide. GeneRuler DNA ladder 100bp and 1kb was used.

Whereas DNA treated with SGM, and SUM was susceptible for TaqI digestion, and two bands were detected on the gel (Figure 2b). To exclude that the observed specificity is due to SNM instability, we tested the degradation of SNM analogs at 65 °C over time. If the degradation is fast, SNM would be unavailable for the reaction. SNM degrade similarly and up to 90% over 60 minutes at 65 °C. However, in the current reaction condition still 100 µM of SNM were available for the reaction compared to the initial concentration (Figure S1). The K_M of SAM for M.Taq is 3.7 µM[36], the amount left in the reaction mixture is well above the required for the methylation reaction.

M.HhaI was observed to accept SGM, SCM, and SUM cofactors for the methylation reaction. dsDNA substrate was protected from digestion with the corresponding restriction enzyme HhaI in the case of SGM, SCM and SUM (Figure 2c). Indeed, the gel showed single band as outcome of the digestion reaction (Figure 2c). The relative efficiency of protection is similar for four different SNM cofactors for M.HhaI by gel. No spontaneous methylation without M.TaqI and M.HhaI was observed using different SNM analogs in these conditions (Figure S2).

Further, we investigated the molecular basis for observed specificity for M.TaqI and promiscuity for M.HhaI by molecular dynamic (MD) simulation. For each representative structure, an all-atom explicit water system was created starting from the relevant crystal structure (PDB IDs 1g38 and 1skm for M.TaqI and M.HhaI, respectively). All ligands, cofactors, and DNA were removed to produce the apo protein structure. To produce the bound structures SNH co-factors were added in alignment with the original bound cofactor. The protein was solvated in a cubic water box padded with 10Å on each side and neutralized with Cl ions. MD simulations were performed using Gromacs 2020v.3[37]. The Amber 14 forcefield[38] was used with TIP3P water[39]. A modified Berendsen thermostat[40] maintained the temperature at 300K and a Parrinello-Rahman barostat[41] maintained a pressure of 1 bar. A 2 fs timestep was employed. Each system was equilibrated for 50 ps in NVT environment followed by 50 ps in NPT environment. The first 10 ns of simulation time were discarded for equilibration purposes. The systems were simulated for an additional 500 ns of production simulation time in an NPT environment with 5 independent replicates. SNH cofactors stayed bound for a minimum of 10 ns indicating a stable bound starting structure. Longer time-scale binding varied between proteins and cofactors. To observe differences in binding patterns, simulations are aligned around the binding pocket defined as alpha carbons of residues

within 7 Å of the crystal structure bound SAH. RMSF of ligand heavy atoms is calculated on the SNH analogs in relation to the aligned structures.

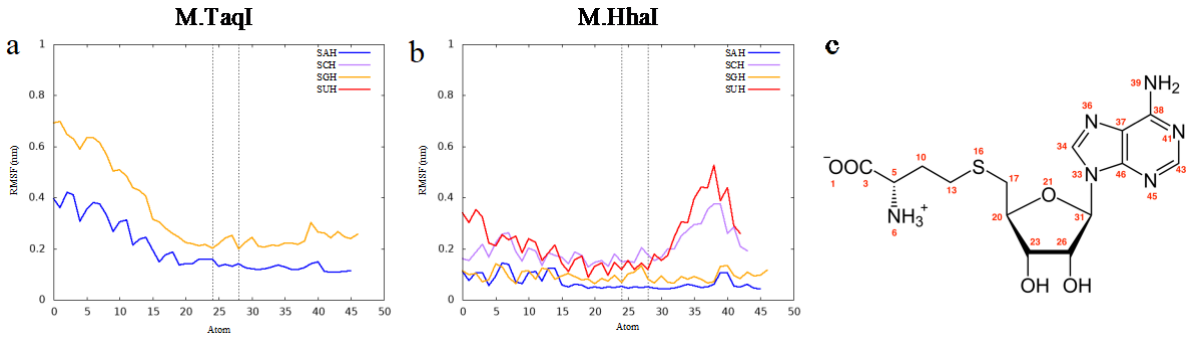

Figure 3: RMSF of heavy atoms on analog SNH cofactors averaged over 500 ns of simulation for (a) M.TaqI and (b) M.HhaI. SAH (blue) and SGH (orange) are shown for M.TaqI and SAH (blue), SCH (purple), SGH (orange), and SUH (red) are shown for M.HhaI. Example atom numbering for SAH is shown in (c); other analogs are numbered similarly starting at the methionine chain. Atoms which participate in the conserved bidentate interactions are denoted by dashed lines on the plots. SCH and SUH simulated with M.TaqI are unbound and not explicitly shown.

M.HhaI and M.TaqI exhibited different cofactor binding patterns as observed by RMSF patterns (Figure 3). M.HhaI tightly binds the methionine chain on each SNH cofactor through multiple backbone interactions within the binding pocket. Most notable of conserved interactions is a backbone hydrogen bonding event between F-18 and -NH3 on the methionine chain. The nucleobase moiety maintains a more flexible position, with SAH and SGH creating nucleobase-mediated hydrogen bonding with D-60 and no analogous hydrogen bonding with SCH or SUH nucleobase moieties. M.TaqI exhibits strong nucleobase binding with SAH, moderate nucleobase binding with SGH, and very poor binding with SUH and SCH. M.TaqI provides high flexibility to the methionine chain, leaving the burden of binding to the nucleobase and ribose moieties.

The binding patterns were further qualitatively exemplified in the length of binding for each SNH analog. In M.HhaI, each of the SNH analogs maintained bound conformations throughout the 500 ns simulations, with the exception of a single replica of SAH binding. This unbinding event is not expected to be indicative of physical phenomena, but rather a result of normal fluctuations and relatively rare events which can happen because of stochasticity in MD simulations. In M.TaqI, SAH and SGH maintained a bound position throughout the 500 ns simulations while SUH and

SCH exhibited very poor binding. SUH unbound from its initial position within 50 ns in each replica. SCH remained bound in only 2 out of the 5 replicas for the complete 500 ns.

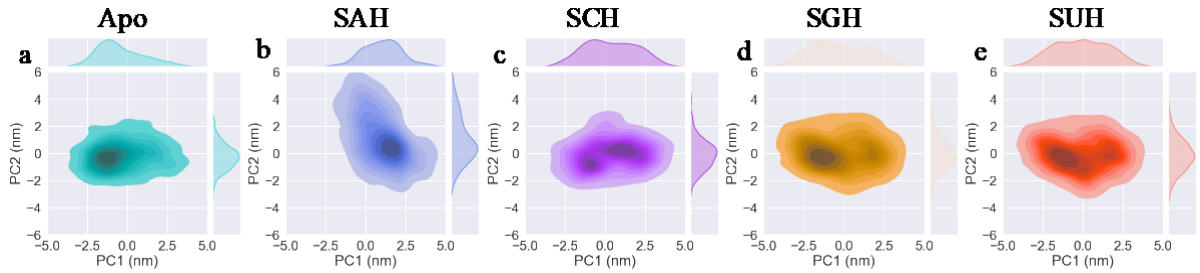

Figure 4: Conformations of 2,500 ns of simulation time plotted on the first and second principal components for M.TaqI in the apo form (a) and bound with SAH (b), SCH (c), SGH (d), and SUH (e). Concentration of color indicates density.

Though binding is conserved through similar bonding interactions in M.TaqI with SAH and SGH analogs, enzymatic activity is not observed with SGM in physical experiments. Protein fluctuations were studied using Principal Component Analysis (PCA) to reduce dimensionality of the conformational space. Dimensionality reduction was performed on protein alpha carbon atoms of 2,500 total ns of simulation time using the Gromacs 2020v3 software package[37] . The resulting populations of protein conformations over the trajectory in the first and second principal components (Figure 4, Figure S4) show varying conformational space covered by SAH in comparison to the apo protein and analog cofactors. SUH, SGH, and apo proteins adopt similar conformations throughout the simulation while SAH adapts to a different conformational space. Based on experimental results, SAH conformations represent the catalytically competent protein conformations. In agreement with experimental results which indicate some efficacy with SCM, M.TaqI bound with SCH maintains the catalytically competent conformational space to some proportion during the simulations.

Our results demonstrate that even though the nucleoside base is far from the catalytic centre it dictates the fate of the methylation reaction. M.TaqI specificity for the SAM is governed by holding the adenosine tightly and high flexibility by methionine chain. Unlike M.TaqI, M.HhaI gives more flexibility to the nucleoside base. Often adenosine is considered as molecular handle, however, our molecular dynamics data indicates that nucleotide base also plays a crucial role in dictating specificity.

Experimentally we observed methylation by M.TaqI using SCH, but simulations showed a poor binding for SCH. This might suggest that tight binding of nucleoside base is necessary for productive binding however dynamics of the enzyme plays an important role as we observed methylation by M.HhaI using all SNH cofactors even after flexibility of the nucleoside base. To see the effect of similar binding interactions for both enzymes in the binding pocket we made mutant I72W M.TaqI. We observed that I72W M.TaqI has same activity as M.TaqI (Figure S3). These data suggest even the exact interacting binding pocket M.TaqI did not change the specificity. Tryptophane gives more ordered and stronger pi-pi interaction compared to isoleucine, but mutant I72W M.TaqI remains specific as nucleoside base plays key in giving the specificity. Finally, we believe that these novel SNM cofactors might be the good candidates for other classes of methyltransferase (RNA, protein, and small molecule methyltransferase) to explore cofactor engineering as synthetic tool to study methyltransferase role in health and disease.

M.G. and P.J. performed methylation experiments. M.G. and P.L. analyzed the methylation data. J.M.G. performed and analyzed molecular dynamics simulation. M.G., P.L. and J.M.G wrote the paper. P.L. conceived and supervised the project.

Acknowledgments

Authors acknowledge to Okinawa Institute of Science and Technology for financial support. This project was supported by a OIST Kick start-up grant. The Laurino lab is supported by a Kakenhi Grant (No. 90812256). We thank Shina Caroline Lynn Kamerlin, Bhanu Pratap Singh Chouhan for scientific discussion.

References

1. K. A. Denessiouk, V. V. Rantanen and M. S. Johnson, *Proteins*, 2001, **44**, 282-291.
2. S. L. Moodie, J. B. Mitchell and J. M. Thornton, *J Mol Biol*, 1996, **263**, 486-500.
3. A. Narunsky, A. Kessel, R. Solan, V. Alva, R. Kolodny and N. Ben-Tal, *Proceedings of the National Academy of Sciences*, 2020, **117**, 4701-4709.
4. P. Rogne, M. Rosselin, C. Grundström, C. Hedberg, U. H. Sauer and M. Wolf-Watz, *Proceedings of the National Academy of Sciences*, 2018, **115**, 3012-3017.

5. I. Nobeli, R. A. Laskowski, W. S. Valdar and J. M. Thornton, *Nucleic Acids Res*, 2001, **29**, 4294-4309.

6. S. Merali and A. B. Clarkson, Jr., *FEMS Microbiol Lett*, 2004, **237**, 179-186.

7. A. E. Pegg, *J Biol Chem*, 2016, **291**, 14904-14912.

8. S. C. Lu, *Mol Aspects Med*, 2009, **30**, 42-59.

9. S. C. Lu and J. M. Mato, *Physiol Rev*, 2012, **92**, 1515-1542.

10. P. K. Chiang, R. K. Gordon, J. Tal, G. C. Zeng, B. P. Doctor, K. Pardhasaradhi and P. P. McCann, *The FASEB Journal*, 1996, **10**, 471-480.

11. M. G. Goll and T. H. Bestor, *Annu Rev Biochem*, 2005, **74**, 481-514.

12. M. Fontecave, M. Atta and E. Mulliez, *Trends Biochem Sci*, 2004, **29**, 243-249.

13. G. L. Cantoni, *Annual Review of Biochemistry*, 1975, **44**, 435-451.

14. B. S. Zhao, I. A. Roundtree and C. He, *Nat Rev Mol Cell Biol*, 2017, **18**, 31-42.

15. G. Zheng, Y. Fu and C. He, *Chemical Reviews*, 2014, **114**, 4602-4620.

16. L. D. Moore, T. Le and G. Fan, *Neuropsychopharmacology*, 2013, **38**, 23-38.

17. R. Holliday and J. E. Pugh, *Science*, 1975, **187**, 226-232.

18. X. Cheng and R. M. Blumenthal, 1999.

19. M. V. C. Greenberg and D. Bourc'his, *Nat Rev Mol Cell Biol*, 2019, **20**, 590-607.

20. N. V. Cornelissen, F. Michailidou, F. Muttach, K. Rau and A. Rentmeister, *Chem Commun (Camb)*, 2020, **56**, 2115-2118.

21. L. Anhauser, F. Muttach and A. Rentmeister, *Chemical Communications*, 2018, **54**, 449-451.

22. C. Dalhoff, G. Lukinavičius, S. Klimašauskas and E. Weinhold, *Nature Chemical Biology*, 2006, **2**, 31-32.

23. G. Lukinavičius, V. Lapienė, Z. Staševskij, C. Dalhoff, E. Weinhold and S. Klimašauskas, *Journal of the American Chemical Society*, 2007, **129**, 2758-2759.

24. G. Lukinavicius, M. Tomkuviene, V. Masevicius and S. Klimasauskas, *ACS Chem Biol*, 2013, **8**, 1134-1139.

25. S. Kim, A. Gottfried, R. R. Lin, T. Dertinger, A. S. Kim, S. Chung, R. A. Colyer, E. Weinhold, S. Weiss and Y. Ebenstein, *Angew Chem Int Ed Engl*, 2012, **51**, 3578-3581.

26. F. Muttach, N. Muthmann, D. Reichert, L. Anhäuser and A. Rentmeister, *Chemical Science*, 2017, **8**, 7947-7953.

27. P. Laurino, A. Toth-Petroczy, R. Meana-Paneda, W. Lin, D. G. Truhlar and D. S. Tawfik, *PLoS Biol*, 2016, **14**, e1002396.

28. G. Basu, D. Sivanesan, T. Kawabata and N. Go, *J Mol Biol*, 2004, **342**, 1053-1066.

29. M. Gade, L. L. Tan, A. M. Damry, M. Sandhu, J. S. Brock, A. Delaney, A. Villar-Briones, C. J. Jackson and P. Laurino, *JACS Au*, 2021, DOI: 10.1021/jacsau.1c00464.

30. K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig and E. Weinhold, *Nat Struct Biol*, 2001, **8**, 121-125.

31. U. T. Sankpal and D. N. Rao, *Crit Rev Biochem Mol Biol*, 2002, **37**, 167-197.

32. G. Schluckebier, M. Kozak, N. Bleimling, E. Weinhold and W. Saenger, *J Mol Biol*, 1997, **265**, 56-67.

33. U. T. Sankpal and D. N. Rao, *Nucleic Acids Res*, 2002, **30**, 2628-2638.

34. T. Naito, K. Kusano and I. Kobayashi, *Science*, 1995, **267**, 897-899.

35. I. Kobayashi, *Nucleic Acids Res*, 2001, **29**, 3742-3756.

36. H. Pues, N. Bleimling, B. Holz, J. Wolcke and E. Weinhold, *Biochemistry*, 1999, **38**, 1426-1434.

37. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem*, 2005, **26**, 1701-1718.

38. V. B. D.A. Case, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman, *AMBER 14*, 2014, University of California, San Francisco.

39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.

40. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684-3690.

41. M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182-7190.