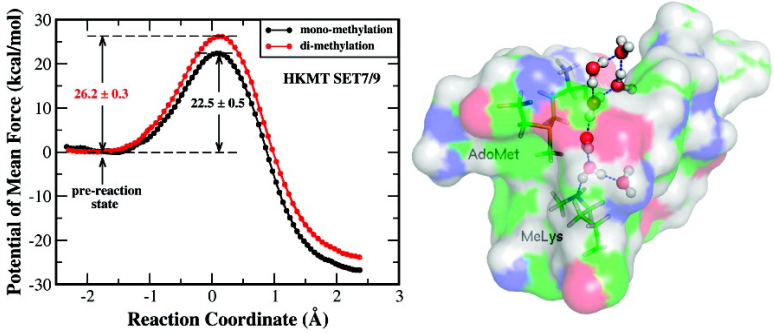
Article

How Do SET-Domain Protein Lysine Methyltransferases Achieve the Methylation State Specificity? Revisited by Ab Initio QM/MM Molecular Dynamics Simulations

Po Hu, Shenglong Wang, and Yingkai Zhang

J. Am. Chem. Soc., **2008**, 130 (12), 3806-3813 • DOI: 10.1021/ja075896n



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 03/01/2008

How Do SET-Domain Protein Lysine Methyltransferases Achieve the Methylation State Specificity? Revisited by Ab Initio QM/MM Molecular Dynamics Simulations

Po Hu, Shenglong Wang, and Yingkai Zhang*

Department of Chemistry, New York University, New York, New York 10003

Received August 6, 2007; E-mail: yingkai.zhang@nyu.edu

Abstract: A distinct protein lysine methyltransferase (PKMT) only transfers a certain number of methyl group(s) to its target lysine residue in spite of the fact that a lysine residue can be either mono-, di-, or tri-methylated. In order to elucidate how such a remarkable product specificity is achieved, we have carried out ab initio quantum mechanical/molecular mechanical (QM/MM) molecular dynamics simulations on two SET-domain PKMTs: SET7/9 and Rubisco large subunit methyltransferase (LSMT). The results indicate that the methylation state specificity is mainly controlled by the methyl-transfer reaction step, and confirm that SET7/9 is a mono-methyltransferase while LSMT has both mono-and di-methylation activities. It is found that the binding of the methylated lysine substrate in the active site of SET7/ 9 opens up the cofactor AdoMet binding channel so that solvent water molecules get access to the active site. This disrupts the catalytic machinery of SET7/9 for the di-methylation reaction, which leads to a higher activation barrier, whereas for the LSMT, its active site is more spacious than that of SET7/9, so that the methylated lysine substrate can be accommodated without interfering with its catalytic power. These detailed insights take account of protein dynamics and are consistent with available experimental results as well as recent theoretical findings regarding the catalytic power of SET7/9.

1. Introduction

Histone lysine methylation plays a pivotal role in regulating chromatin structure and gene expression.[1−5] The enzymes responsible for this essential post-translational modification are protein lysine methyltransferases (PKMTs),[6−13] which catalyze the transfer of methyl group(s) from the cofactor *S*-adenosylmethionine (AdoMet) to some specific lysine residues. In spite of the fact that most PKMTs share a common structural fold—the SET-domain[6,9]—and that a lysine residue can be either mono-, di-, or tri-methylated, a specific protein lysine methyltransferase (PKMT) only transfers a certain number of methyl group(s) to its target lysine residue.[14−23] Recent experimental studies have begun to show that distinct methylation states give rise to different functional consequences.[4,6,24−26] For instance, HP1 chromodomain selectively recognizes di- and tri-methylated H3−K9 but does not bind with the mono-methylated or unmodified target,[24,25] and the binding affinity of ING2 PHD finger dramatically decreases with respect to the decrement of number of methyl groups on H3−K4.[26] Thus, it is of great interest and fundamental importance to elucidate how such a remarkable methylation state specificity is achieved by PKMTs in the first place.

Among SET-domain PKMTs, SET7/9 and Rubisco large subunit methyltransferase (LSMT) are two of the best character-

(1) Kouzarides, T. *Cell* **2007**, *128*, 693−705.
(2) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838−849.
(3) Biel, M.; Wascholowski, V.; Giannis, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 3186−3216.
(4) Sims, R. J.; Nishioka, K.; Reinberg, D. *Trends Genet.* **2003**, *19*, 629−639.
(5) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41−45.
(6) Xiao, B.; Wilson, J. R.; Gamblin, S. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 699−705.
(7) Blackburn, G. M.; Gamblin, S. J.; Wilson, J. R. *Helv. Chim. Acta* **2003**, *86*, 4000−4006.
(8) Cheng, X.; Collins, R. E.; Zhang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267−294.
(9) Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X. *Genome Biol.* **2005**, *6*, 227.
(10) Qian, C.; Zhou, M. M. *Cell. Mol. Life Sci.* **2006**, *63*, 2755−2763.
(11) Couture, J. F.; Trievel, R. C. *Curr. Opin. Struct. Biol.* **2006**, *16*, 753−760.
(12) Min, J.; Zhang, X.; Cheng, X.; Grewal, S. I.; Xu, R. M. *Nat. Struct. Biol.* **2002**, *9*, 828−832.
(13) Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. M. *Cell* **2003**, *112*, 711−723.
(14) Wilson, J. R.; Jing, C.; Walker, P. A.; Martin, S. R.; Howell, S. A.; Blackburn, G. M.; Gamblin, S. J.; Xiao, B. *Cell* **2002**, *111*, 105−115.
(15) Trievel, R. C.; Beach, B. M.; Dirk, L. M. A.; Houtz, R. L.; Hurley, J. H. *Cell* **2002**, *111*, 91−103.
(16) Zhang, X.; Tamaru, H.; Khan, S. I.; Horton, J. R.; Keefe, L. J.; Selker, E. U.; Cheng, X. *Cell* **2002**, *111*, 117−127.
(17) Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, G. M.; Gamblin, S. J. *Nature* **2003**, *421*, 652−656.
(18) Trievel, R. C.; Flynn, E. M.; Houtz, R. L.; Hurley, J. H. *Nat. Struct. Biol.* **2003**, *10*, 545−552.
(19) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12*, 177−185.
(20) Manzur, K. L.; Farooq, A.; Zeng, L.; Plotnikova, O.; Koch, A. W.; Zhou, M. M. *Nat. Struct. Biol.* **2003**, *10*, 187−196.
(21) Xiao, B.; Jing, C.; Kelly, G.; Walker, P. A.; Muskett, F. W.; Frenkiel, T. A.; Martin, S. R.; Sarma, K.; Reinberg, D.; Gamblin, S. J.; Wilson, J. R. *Genes Dev.* **2005**, *19*, 1444−1454.
(22) Couture, J. F.; Collazo, E.; Brunzelle, J. S.; Trievel, R. C. *Genes Dev.* **2005**, *19*, 1455−1465.
(23) Qian, C.; Wang, X.; Manzur, K.; Farooq, A.; Zeng, L.; Wang, R.; Zhou, M. M. *J. Mol. Biol.* **2006**, *359*, 86−96.
(24) Jacobs, S. A.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080−2083.
(25) Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. *Nature* **2002**, *416*, 103−107.
(26) Pena, P. V.; Davrazou, F.; Shi, X.; Walter, K. L.; Verkhusha, V. V.; Gozani, O.; Zhao, R.; Kutateladze, T. G. *Nature* **2006**, *442*, 100−103.

10.1021/ja075896n CCC: $40.75

ized both experimentally[14,15,17,18,27−30] and theoretically.[31−34] SET7/9 is a mono-methyltransferase that only catalyzes the transfer of one methyl group to the target lysine residue,[14,17] whereas LSMT is a tri-PKMT that can methylate the target lysine residue all the way to terminal product tri-methylated lysine.[15,18,29] For both enzymes, the cofactor AdoMet and the substrate peptide bind to opposite faces of the SET-domain and are connected by a narrow channel that has a hydrophobic inner wall,[17,18,27,29] and the target lysine residue is inserted into this narrow channel to access the methyl moiety of AdoMet. Alternatively, the geometry and shape of the bottom of the lysine access channel, which binds the ϵ-amino group of the substrate, are found to be quite different. This ϵ-amino group binding pocket is found to be more spacious in the tri-methylase LSMT than that in the mono-methylase SET7/9.[17,18] Thus, it is quite intuitively appealing to postulate that multi-methylation activity is dependent on the ability of accommodating the increasing bulk of the lysine ϵ-amino group in the active site. Such a steric hindrance hypothesis seems to be be well supported by relevant mutation studies, which have shown that altering the size of a key residue in this binding pocket can change the methylation state specificity of PKMTs.[8,21−23,28] However, it arises a fundamental and intriguing question: how can such a steric effect be imposed given the fact that a protein structure in solution is very dynamic in nature and the methyl group is very small? For SET7/9 to be lack of di-methylation activity, is it due to the disruption of the formation of the near-attack reactive conformation, or the destabilization of the transition state, or the steric hindrance to accommodate the methylated product, or other factors? Although previous studies suggested that the formation of near-attack conformations (NAC) could be an important factor,[18,31,34] the evidence is qualitative in nature but not conclusive. Meanwhile, our recent ab initio QM/MM molecular dynamics simulation studies on the mono-methylation reaction in SET7/9 have shown that the formation of NAC is not a major source of its catalytic power.[32] Thus, it is clear that some important questions regarding the mechanism controlling methylation state specificity of PKMTs remain to be addressed.

In this work, we have carried out ab initio QM/MM molecular dynamics simulations[32,35−38] on the di-methylation reaction in SET7/9, and both mono- and di-methylations catalyzed by Rubisco large subunit methyltransferase (LSMT). These simulations together with our previous calculation on the mono-methylation reaction in SET7/9 have provided detailed insights into nature's solution to achieve the product specificity of lysine methylation.

2. Methods

To investigate chemical reactions in complex systems, statistical sampling on a reasonably accurate potential energy surface is needed to obtain reliable results. Here, our theoretical approaches center on Born−Oppenheimer MD simulations with ab initio QM/MM potential[39−48] and the umbrella sampling method.[49−51] At each time step, the atomic forces as well as the total energy of the QM/MM system are calculated with a pseudobond ab initio QM/MM approach[47,52−54] on-the-fly, and Newton equations of motion are integrated. From a series of biased simulations, the potential of mean force along the reaction coordinate is obtained with the weighted histogram analysis method (WHAM).[55−57] This direct ab initio QM/MM MD approach, which takes account of dynamics of reaction active site and its environment on an equal footing, has been recently demonstrated to be feasible and successful in elucidating the catalytic power of SET7/9.[32] The calculated free energy profiles for the lysine mono-methylation catalyzed by SET7/9, and its corresponding uncatalyzed reaction in aqueous solution are very consistent with the available experimental results. In our current study, a very similar computational protocol has been employed to investigate the di-methylation reaction in SET7/9, as well as for the mono- and di-methylations catalyzed by LSMT.

There are four simulation systems: SET7/9-AdoMet-Lys, SET7/9-AdoMet-MeLys, LSMT-AdoMet-Lys, and LSMT-AdoMet-MeLys, which are enzyme−substrate complexes for mono- and di-methylation reactions in SET7/9 and LSMT respectively. Both SET7/9-AdoMet-Lys and SET7/9-AdoMet-MeLys systems were prepared on the basis of the crystal structure 1O9S,[17] which has been described in detail previously.[31] The initial structure of the reactant complex for the mono-methylation was constructed by moving the methyl group from methyl lysine residue to the sulfur atom of AdoHcy. In preparation for the initial structure for the di-methylation reaction, the hydrogen atom which is close to Tyr305 was replaced with a methyl group and the N−C bond distance was changed to the same as that in the crystal structure 1O9S.[17] The rationale that we replace the hydrogen atom close to Tyr305 instead of the one hydrogen-bonded to Tyr245 is that the hydrogen bond between lysine and Tyr305 is much weaker than that between lysine and Tyr245 based on the crystal structure 1O9S as well as our simulations on the reactant complex for the mono-methylation. The crystal structure 1P0Y,[18] a ternary complex of Rubisco LSMT with *S*-adenosyl-homocysteine (AdoHcy) and a methylated lysine residue, has been employed to prepare both LSMT-AdoMet-Lys and LSMT-AdoMet-MeLys systems. Considering that there is a crystal water molecule close to the N_ξ atom of the lysine in the structure 1OZV[18] and the corresponding position in the structure of 1P0Y[18] is empty, a water molecule is placed into the enzyme active site of 1P0Y by superimposing the structures of 1P0Y and 1OZV.[18] Missing hydrogen

(27) Kwon, T.; Chang, J. H.; Kwak, E.; Lee, C. W.; Joachimiak, A.; Kim, Y. C.; Lee, J. W.; Cho, Y. *EMBO J.* **2003**, *22*, 292−303.
(28) Collins, R. E.; Tachibana, M.; Tamaru, H.; Smith, K. M.; Jia, D.; Zhang, X.; Selker, E. U.; Shinkai, Y.; Cheng, X. *J. Biol. Chem.* **2005**, *280*, 5563−5570.
(29) Couture, J. F.; Hauk, G.; Thompson, M. J.; Blackburn, G. M.; Trievel, R. C. *J. Biol. Chem.* **2006**, *281*, 19280−19287.
(30) Couture, J. F.; Collazo, E.; Hauk, G.; Trievel, R. C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 140−146.
(31) Hu, P.; Zhang, Y. *J. Am. Chem. Soc.* **2006**, *128*, 1272−1278.
(32) Wang, S.; Hu, P.; Zhang, Y. *J. Phys. Chem. B* **2007**, *111*, 3758−3764.
(33) Zhang, X.; Bruice, T. C. *Biochemistry* **2007**, *46*, 5505−5514.
(34) Guo, H. B.; Guo, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8797−8802.
(35) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1993**, *97*, 11868.
(36) Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Rohrig, U.; Spiegel, K.; Sulpizi, M.; Vandevondele, J.; Zumstein, M.; Rothlisberger, U. *Chimia* **2002**, *56*, 13−19.
(37) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. *J. Phys. Chem. B* **2004**, *108*, 4210−4220.
(38) Crespo, A.; Marti, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Am. Chem. Soc.* **2005**, *127*, 6940−6941.
(39) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(40) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718−730.
(41) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(42) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(43) Cui, Q.; Karplus, M. *Adv. Protein Chem.* **2003**, *66*, 315−372.
(44) Shurki, A.; Warshel, A. *Adv. Protein Chem.* **2003**, *66*, 249−313.
(45) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.
(46) Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393−1402.
(47) Zhang, Y. *Theor. Chem. Acc.* **2006**, *116*, 43−50.
(48) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173−290.
(49) Patey, G. N.; Valleau, J. P. *J. Chem. Phys.* **1975**, *63*, 2334−2339.
(50) Boczko, E. M.; Brooks, C. L. *J. Phys. Chem.* **1993**, *97*, 4509−4513.
(51) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(52) Zhang, Y.; Lee, T. S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46−54.
(53) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483−3492.
(54) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
(55) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(56) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(57) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40−57.

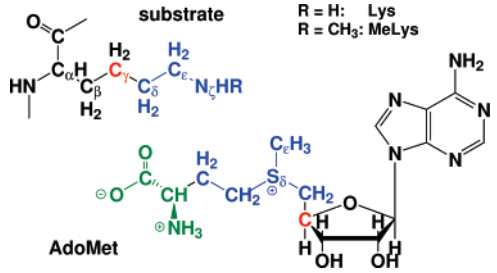


Figure 1. Illustration of the division of the QM/MM system for simulating the methyl-transfer from AdoMet to the substrate Lys or MeLys. The HF-(6-31G*) method was employed to model the reaction center (colored in blue), and the 3-21G* basis set was used for atoms shown in green. Two boundary carbon atoms (red) were treated with improved pseudobond parameters.[54] All other atoms belong to the MM subsystem.

atoms were added via the program pdb2pqr,[58] and each enzyme−substrate complex was solvated with a 27.0 Å solvent water sphere centered on the active site (the sulfur atom of AdoMet). To take account of possible structural reorganizations and protein dynamics, each prepared system was equilibrated with minimizations and 100 ps molecular dynamics simulations. The resulting snapshot was further minimized by the MM method, and then an iterative minimization procedure with the reaction coordinate driving method has been employed to map out a minimum energy path with ab initio QM/MM calculations. For each determined structure along the path, the MM subsystem was further equilibrated by carrying out 500 ps MD simulations with the MM force field. Finally, the resulting snapshot was used as the starting structure for ab initio QM/ MM MD simulations in that specific umbrella window.

The partition of the QM/MM system is illustrated in Figure 1 and the QM/MM interface has been described by a pseudobond approach.[52,54] Since only $S_\delta\cdots C_\epsilon$ and $N_\xi\cdots C_\epsilon$ bonds directly participate in the methyl-transfer reaction, the reaction coordinate is defined as the difference of two bond lengths: $RC = d_{S_\delta\cdots C_\epsilon} - d_{N_\xi\cdots C_\epsilon}$. HF(6-31G*) method, which has been well-known to describe such methyl-transfer reactions very well with a reasonable computational cost,[59−64] was employed to model the reaction center (colored in blue as shown in Figure 1). In order to minimize the computational cost, 3-21G* basis set was used for atoms colored in green, which do not directly participate in the methyl-transfer reaction. Two boundary carbon atoms (colored in red) were treated with improved pseudobond parameters.[54] There are 34 QM atoms for the mono-methylation reaction system (SET7/9-AdoMet-Lys and LSMT-AdoMet-Lys), and 37 QM atoms for simulating both SET7/9-AdoMet-MeLys and LSMT-AdoMet-MeLys. All other atoms in AdoMet, histone peptide, and enzyme were described with the Amber molecular mechanical force field,[65,66] and the TIP3P model[67] has been employed for water molecules. Such a small QM/MM partition and the employment of HF(6-31G*/3-21G*) QM/MM method has been found to be able to describe lysine methylation reaction well in our previous studies,[32] and it overestimates the reaction barrier about 3 kcal/mol in comparison with MP2(6-31+G*) QM/MM calculations with a large QM subsystem (66 atoms).[31]

In ab initio QM/MM MD simulations with the umbrella sampling method,[49−51] the total potential energy of the system was biased with a harmonic potential, centered on successive values of the reaction coordinate. The forces on atoms in both QM and MM subsystems as well as the total energy are calculated on-the-fly with the QM/MM method at each time step (1 fs), and Newton equations of motion are integrated with Beeman algorithm.[68] For each methyl-transfer reaction, we have employed about 35 umbrella windows with harmonic potential force constants 40−60 kcal·mol^{-1}·Å^{-2}. For each umbrella window, water molecules in a sphere of 23.0 Å are retained and all atoms within 20.0 Å sphere of sulfur atom of AdoMet have been sampled for 30 ps. The configurations were recorded every 10 steps (10 fs) and collected for 20 ps for the data analysis after an equilibration period of 10 ps. Thus, the total ab initio QM/MM MD simulation time length for each enzyme reaction is around 1 ns. The probability distributions (e.g., histograms) along the reaction coordinate were determined for each window and pieced together with the weighted histogram analysis method (WHAM)[55−57] to calculate the potential of mean force.

All of our simulations have been carried out with modified version of Gaussian 03,[69] Q-Chem,[70] and TINKER[71] programs. The spherical boundary condition has been applied so that atoms outside of 20.0 Å of the sulfur atom of the AdoMet are fixed. A cutoff of 12.0 Å was used for van der Waals interactions, and a cutoff of 18.0 Å was employed for electrostatic interactions among MM atoms. There is no cutoff for electrostatic interactions between QM and MM atoms. Berendsen thermostat method[72] has been used to control the system temperature at 300 K.

3. Results and Discussions

3.1. Free Energy Reaction Profiles and Transition State Geometries. By employing on-the-fly ab initio QM/MM MD simulations with umbrella sampling method, we have computed potentials of mean force (PMF) for the di-methylation in SET7/9, mono- and di-methylations in LSMT, as shown in Figure 2. The PMF for the mono-methylation in SET7/9, which has been determined previously,[32] is also presented for comparison. We can see that all free energy reaction profiles converge well, and the difference between the PMF curves obtained from different time periods (10−20 ps versus 20−30 ps) is quite small. For all four PMF curves, they are quite flat and near the minimum around the reaction coordinate of −1.7 to −1.9 Å. To be consistent with our previous study, we have also chosen the same reaction coordinate of −1.75 Å as a pre-reaction state. For the mono-methylation reaction catalyzed by SET7/9, the calculated free energy barrier is 22.5 ± 0.5 kcal/mol,[32] which is in excellent agreement with the activation barrier of 20.9 kcal/mol estimated from the experimental k_{cat} value of 0.004 s^{-1}.[15] Our calculated reaction barrier for the di-methylation reaction in SET7/9 is 26.2 ± 0.3 kcal/mol, which indicates that the di-methylation reaction would be about 500-fold slower than the mono-methylation reaction. Considering that the measured k_{cat} for the mono-methylation reaction catalyzed by SET7/9 is already quite small, the 500-fold rate reduction would lead to no detectable activities

(58) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665−W667.
(59) Bento, A. P.; Sola, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1497−1504.
(60) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049−3050.
(61) Morokuma, K. *J. Am. Chem. Soc.* **1982**, *104*, 3732−3733.
(62) Truong, T. N.; Stefanovich, E. V. *J. Phys. Chem.* **1995**, *99*, 14700−14706.
(63) Hase, W. L. *Science* **1994**, *266*, 998−1002.
(64) Mohamed, A. A.; Jensen, F. *J. Phys. Chem. A* **2001**, *105*, 3259−3268.
(65) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(66) Markham, G. D.; Norrby, P. O.; Bock, C. W. *Biochemistry* **2002**, *41*, 7636−7646.
(67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(68) Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130−139.
(69) Frisch, M. J.; et al. *Gaussian 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.
(70) Shao, Y.; et al. *Q-Chem*, Version 3.0; Q-Chem, Inc.: Pittsburgh, PA, 2006.
(71) Ponder, J. W. *TINKER, Software Tools for Molecular Design, Version 4.2*. The most updated version for the TINKER program can be obtained from J. W. Ponder's World Wide Web site at http://dasher.wustl.edu/tinker., Jun. 2004.
(72) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(73) DeLano, W. L. The PyMOL Molecular Graphics System, 2002. DeLano Scientific, Palo Alto, CA, U.S.A.

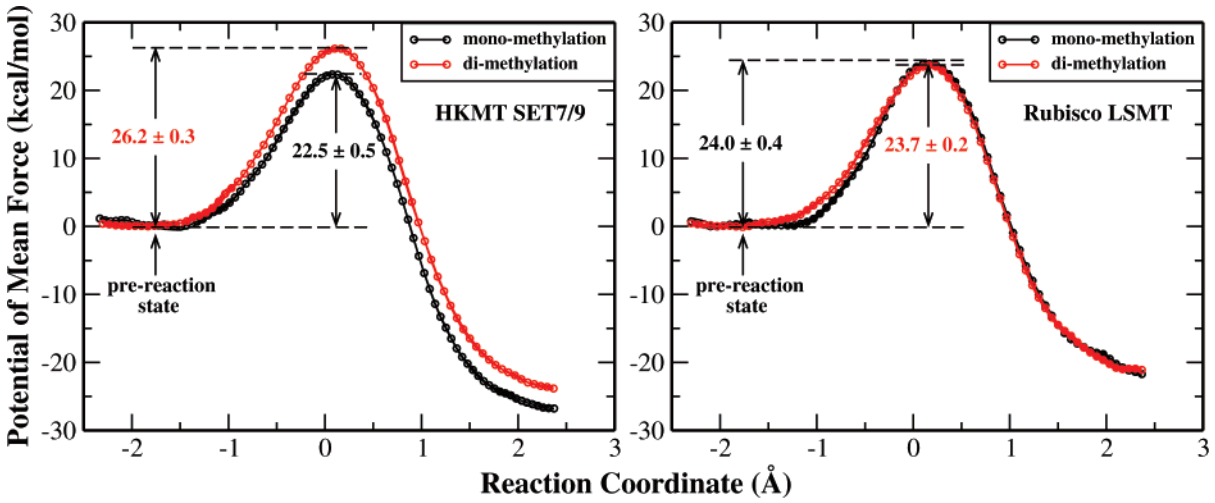


Figure 2. Potentials of mean force (PMF) for mono- and di-methylations catalyzed by SET7/9 and LSMT, respectively. The left side corresponds to the reactant and the right side is for the product. For each reaction, three PMF curves from different simulation time intervals (10−30 ps, 10−20 ps, and 20−30 ps) have been plotted.

for the di-methylation reaction in SET7/9. Thus, our calculation results here are very consistent with the experimental results that HKMT SET7/9 is exclusively a mono-methyltransferase which is unable to catalyze di-methylation reaction.[14,17] For the Rubisco LSMT catalyzed methylations, the free energy of activation are calculated to be 24.0 ± 0.4 and 23.7 ± 0.2 kcal/mol for mono- and di-methylation respectively, in excellent agreement with the values of 23.3 and 22.5 kcal/mol estimated from experimentally determined k_{cat} values.[18] Thus, our results here confirm that SET7/9 is a mono-methyltransferase, whereas LSMT has both mono- and di-methylation activities, and indicate that the methyl-transfer reaction step is an important step in controlling the methylation state specificity.

For the calculated potentials of mean force in the reactant region as shown in Figure 2, we can see that there is clearly a minimum well associated with the reaction coordinate from −1.5 to −1.9 Å for the mono-methylation in SET7/9 (the difference in PMF between RC = −1.7 and RC = −2.5 is about 2 kcal/mol), whereas the curves are very flat for the formation of the pre-reaction states for other three simulation systems (the differences in PMF between RC = −1.7 and RC = −2.5 are all below 0.5 kcal/mol). This indicates that the substrate binding in SET7/9 mono-methylation is stronger than that in SET7/9 di-methylation, LSMT mono- and di-methylations. Meanwhile, these results suggest that the formation of near-attack reactant conformers is not likely to be the controlling factor in determining the methylation state specificity.

Since the di-methylation reaction in SET7/9, mono- and di-methylation reactions in LSMT are characterized by ab initio QM/MM MD simulations for the first time, it would be interesting to characterize their transitions states. From Figure 2, we can see that the locations of their transition states are very consistent, which are around the reaction coordinate of 0.1 Å. The calculated average key geometry values are shown in Table 1. These results further confirm the in-line nucleophilic substitution mechanism and a mainly dissociate transition state for enzyme catalyzed lysine methylation reactions, and are consistent with the values obtained with B3LYP(6-31G*) QM/MM[31] and SCC-DFTB QM/MM[33,34] methods. Meanwhile, Table 1 also indicates that the product specificity of protein lysine methylation does not originate from the nature of the transition state.

Table 1. Comparison of Transition State Geometries of HKMT SET7/9 and Rubisco LSMT Mono- and Di-methylations[a]

	HKMT SET7/9		Rubisco LSMT	
	mono-methylation	di-methylation	mono-methylation	di-methylation
$S_\delta \cdots C_\epsilon$	2.35 ± 0.06	2.35 ± 0.05	2.35 ± 0.05	2.35 ± 0.05
$N_\zeta \cdots C_\epsilon$	2.25 ± 0.06	2.25 ± 0.05	2.20 ± 0.05	2.20 ± 0.05
$S_\delta \cdots C_\epsilon \cdots N_\zeta$	171.4 ± 4.0	172.7 ± 3.6	173.3 ± 3.9	174.0 ± 3.1

[a] Bond lengths in Å and angle in deg.

3.2. Origin of Methylation State Specificity. Our calculated results so far indicate that the methyl-transfer step is an important step in controlling the methylation state specificity. The key remaining question is why the free energy difference between the pre-reaction state (RC = −1.75 Å) and the transition state is significantly higher in di-methylation reaction in SET7/9 than in its corresponding mono-methylation reaction, whereas the reaction barriers are very comparable with each other for mono- and di-methylations in LSMT. Here, we have analyzed the electrostatic interaction energies between the QM subsystem and its enzyme environment for both pre-reaction and transition states with all four simulated reactions, as shown in Figures 3 and 4. In order to provide more detailed insights, the total effect of the enzyme environment has been divided into specific contributions of the protein and water molecules, respectively. It should be noted that the contribution of water molecules refers to all water molecules in the enzyme system. For SET7/9, two left panels in Figure 3 indicate that the shift in protein contribution to the electrostatic interaction energy from the pre-reaction state to the transition state is quite similar. In both mono- and di-methylation reactions, the interaction energy is more negative in the transition state than that in the pre-reaction state, which indicates that the protein environment stabilizes the transition state. However, there is a clear difference in the contribution from water molecules, as shown in two right panels of Figure 3. For the mono-methylation in SET7/9, the distribution curve of contribution from water molecules only slightly shifts to the positive side from the pre-reaction state to the transition state, whereas such a shift is much larger for the di-methylation reaction. These results indicate that water

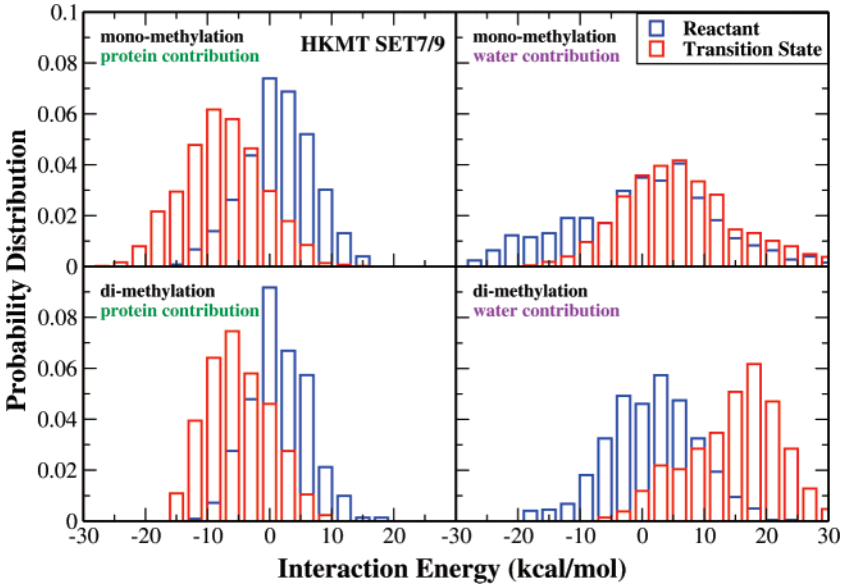


Figure 3. Calculated electrostatic interaction energies between the QM subsystem and its environment (protein or water) for both pre-reaction and transition states in SET7/9 mono- and di-methylation. In each panel, the average of energy contribution at the pre-reaction state (blue) has been shifted to be zero and the same value is shifted for that at the corresponding transition state.

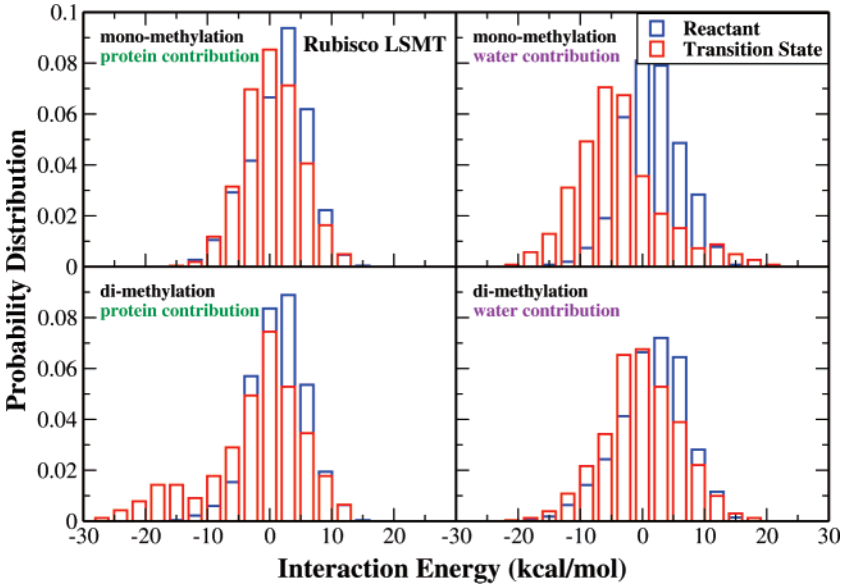


Figure 4. Calculated electrostatic interaction energies between the QM subsystem and its environment (protein or water) for both pre-reaction and transition states in LSMT mono- and di-methylation. In each panel, the average of energy contribution at the pre-reaction state (blue) has been shifted to be zero, and the same value is shifted for that at the corresponding transition state.

molecules in SET7/9 enzyme system are more unfavorable for the di-methylation reaction, which leads to the higher reaction barrier than the mono-methylation reaction. Alternatively, for LSMT as seen from Figure 4, the shift from the pre-reaction state to the transition state are very similar between mono- and di-methylations for both protein and solvent molecules contributions. These results clearly demonstrate that a key factor in determining the free energy difference between mono- and di-methylation comes from the contribution of water molecules in the enzyme system.

In order to further understand the structural origin of this interaction energy difference, we have calculated the distribution of water molecules in the enzyme active site for both pre-reaction and transition states in all four simulation systems, as shown in Figures 5 and 6. We can see that for the pre-reaction states, the occupancy of water molecules near N_ζ is very different between mono- and di-methylation reactions in SET7/9, whereas it is quite similar between two methylations in LSMT. Meanwhile, from the pre-reaction state to the transition state, the distribution of water molecules near N_ζ has a large shift for the di-methylation reaction in SET7/9, whereas there is little change in other three simulation systems. These results indicate that the higher barrier for the di-methylation in SET7/9 than other simulated reaction systems may come from the different arrangement of water molecules in the enzyme active site.

By examining the snapshots from ab initio QM/MM MD simulations, a distinct feature in the pre-reaction state of the di-methylation reaction in SET7/9 is found: the lone pair of N_ζ atom of the substrate lysine residue often forms a direct

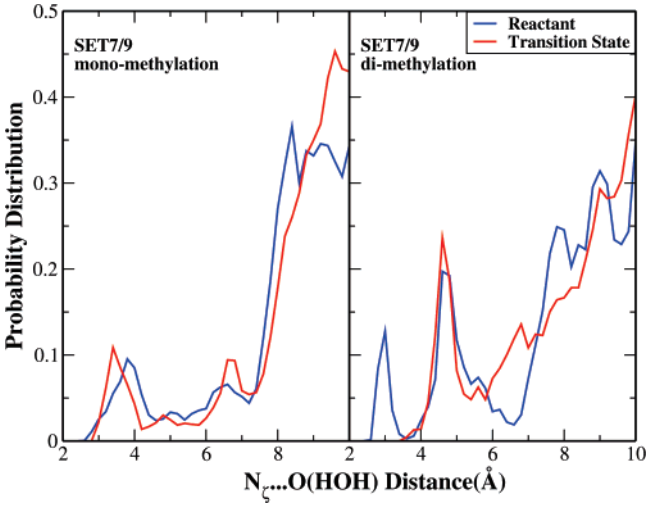


Figure 5. Distribution of water molecules (oxygen atom) near the N_ζ atom of the substrate Lys or MeLys for the pre-reaction states in SET7/9 mono- and di-methylations.

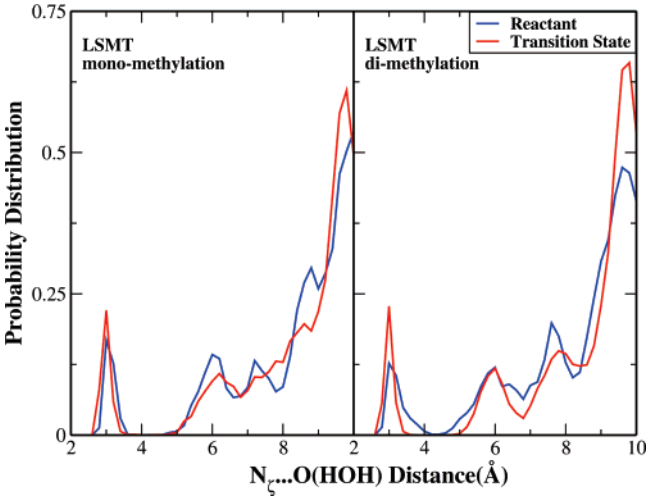


Figure 6. Distribution of water molecules (oxygen atom) near the N_ζ atom of the substrate Lys or MeLys for the pre-reaction states in LSMT mono- and di-methylations.

N···H−O hydrogen bond with a water molecule, which is connected with solvent water molecules on the protein surface through a water chain, as illustrated in Figure 7. This water chain usually contains about four water molecules and is located in the side of cofactor AdoMet binding channel. Since the lone pair of N_ζ is required to nucleophilic attack the methyl group during the reaction, such an N···H−O hydrogen bond needs to be broken for the methyl-transfer. During the di-methylation reaction in SET7/9, the average shortest N_ζ···H(HOH) distance has changed from 2.23 Å at the pre-reaction state to 4.64 Å at the transition state, which is accompanied by the reduction of an average of 0.8 water molecule in the channel. Meanwhile, our simulation results indicate that such an N···H−O hydrogen bond as well as the water chain do not exist in other three simulation systems. In the case of the mono-methylation reaction in SET7/9, there is more or less a similar channel, but the channel radius is smaller. The calculated average minimum channel radii for mono- and di-methylation are 1.42 ± 0.14 and 1.61 ± 0.12 Å via the hole program,[76] respectively. The smaller size of the channel in the mono-methylation does not allow solvent water molecules to get in, as shown is Figure 8.

In LSMT mono- and di-methylation, the water molecule closest to the N_ζ atom corresponds to the crystal water molecule in the active site of the crystal structure 1OZV,[18] and the lone pair of oxygen atom of water molecule forms an N−H···O hydrogen bond with substrate lysine residue. Such an N−H···O hydrogen bond, which is clearly very different from the N···H−O hydrogen bond, actually becomes even a little stronger during the reaction and facilitates the methyl-transfer to some extent. Thus, comparing with either mono-methylation in SET7/9 or methylations in LSMT, extra energy has been paid to break the N···H−O hydrogen bond as well as to push a water molecule out of the channel during the di-methylation reaction in SET7/9, which results in a higher reaction activation barrier.

It should be noted that our results here are very different from the hypothesis that has been put forward by Bruice and his co-workers[33,74,75] regarding the product specificity of histone lysine methylation. They have suggested that whether a certain step of methyl-transfer reaction can happen or not depends on the ability of lysine/methyl-lysine deprotonation, and the deprotonation step is dependent on the formation of a water chain in the complex of enzyme and protonated lysine substrate. However, the deprotonation step is not the rate-determining step, and its calculated barrier for proton transfer is ∼8.4 kcal/mol,[75] which is more than 10 kcal/ mol lower than that for the methyl-transfer step. In addition, as experimental biochemical studies indicated that SET7/9 and DIM-5 are active at pH = 8 or higher and have an unusual high pH optimum of ∼10,[14,16] it is possible that the substrate Lys/MeLys side chain is deprotonated before it binds to SET-domain containing PKMTs. These may cast some doubts on the hypothesis that the lysine/methyl-lysine deprotonation is essential for the methylation state specificity. Here, our results emphasize the rate-determining methyl-transfer step itself rather than the deprotonation step. We find that in the complex of SET7/9, AdoMet and the deprotonated MeLys substrate, the access of solvent water molecules to the enzyme active site through a water chain results in a higher activation barrier for the di-methylation reaction in SET7/9.

Our new insights regarding the methylation state specificity are very consistent with our current understanding of the catalytic power of SET7/9. In previous simulations, it has been found that the enzyme SET7/9 lowers the reaction activation barrier for the lysine mono-methylation step by 8.4 kcal/mol comparing with the corresponding reaction in aqueous solution.[32] An essential contributor to this significant reaction rate enhancement is due to a combination of electrostatic pre-organization in enzyme and the hydrogen bond network reorganization in solution. For the uncatalyzed lysine mono-methylation reaction in aqueous solution, one key finding is that the water distribution around the lysine nitrogen undergoes a significant change from the pre-reaction state to the transition state.[32] In the pre-reaction state of the solution reaction, the N···H−O hydrogen bond is formed between the solvent water molecule and the lone-pair electron of lysine nitrogen, which has to be broken during the reaction process. It has been found that the water molecules close to the lysine nitrogen atom are very unfavorable for the methyl-transfer reaction in the aqueous solution. So we can see that the di-methylation reaction in SET7/9 shares some similar

(74) Zhang, X.; Bruice, T. C. *Biochemistry* **2007**, *46*, 9743−9751.
(75) Zhang, X.; Bruice, T. C. *Biochemistry* **2007**, *46*, 14838−14844.
(76) Smart O. S.; Goodfellow, J. M.; Wallace, B. A. *Biophys. J.* **1993**, *65*, 2455−2460.

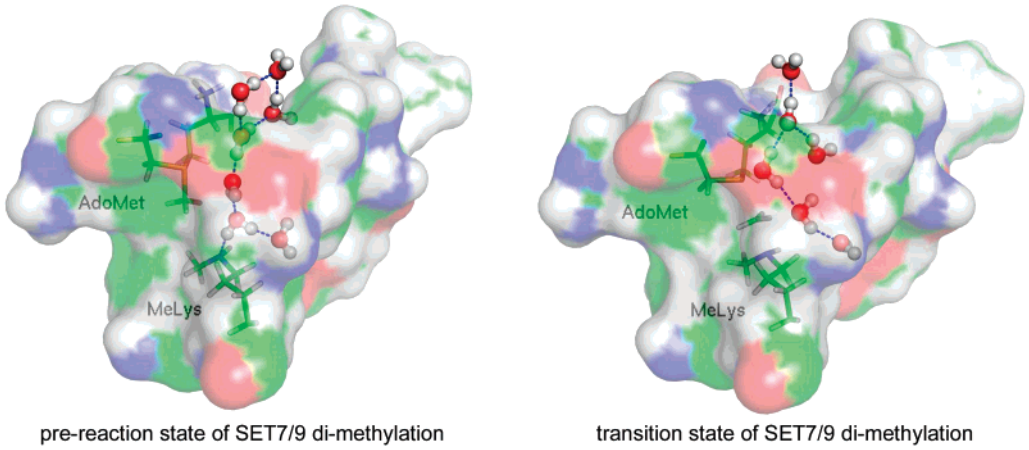


Figure 7. Illustration of a water chain connecting substrate lysine and surface water in the pre-reaction state and transition state of HKMT SET7/9 di-methylation, respectively. Figures are rendered in PyMOL.[73]

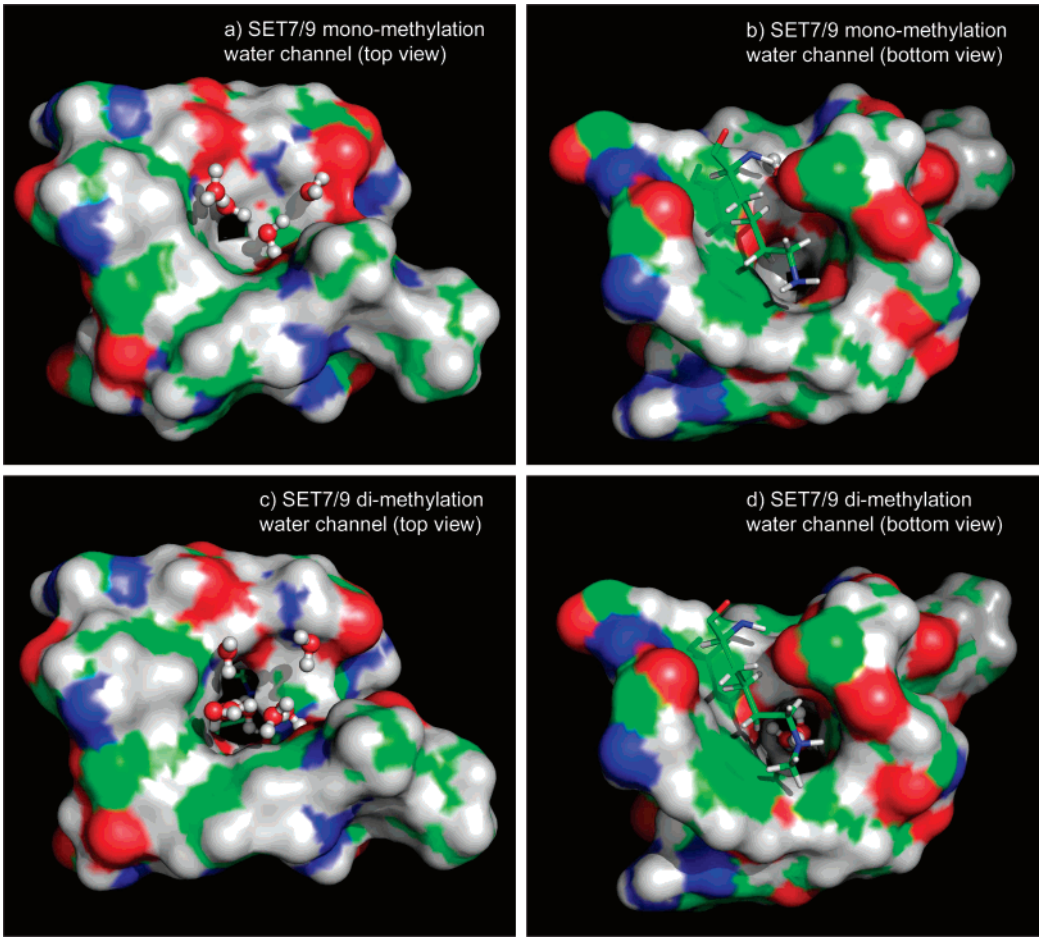


Figure 8. Illustration of the channel in HKMT SET7/9 mono- and di-methylation, respectively. (a) channel in mono-methylation (view from top); (b) channel in mono-methylation (view from bottom); (c) channel in di-methylation (view from top); (d) channel in di-methylation (view from bottom). The water molecules can get access into the channel in the di-methylation, but it is not the case in the mono-methylation.

features with the lysine methylation reaction in aqueous solution, which results in the higher reaction barrier and lower methylation activity.

From the above results and discussions, a dynamic mechanism is emerged regarding how the methylation state specificity is achieved by SET-domain PKMTs. The mono-methylase SET7/9 generally has a quite tight binding pocket in the mono-methylation process, as demonstrated clearly in the determined structures and simulation studies. However, the main controlling factor for its lacking of di-methylation activity is not the steric hindrance which results in the disruption of the formation of the near-attack reactive conformation. Due to the dynamic nature of the protein structure in solution, the binding pocket of SET7/9 can be enlarged to accommodate the additional methyl group in its active site during the di-methylation reaction, and it also opens up the binding channel of AdoMet a little bit. The consequence is that solvent water molecules are allowed to penetrate in the channel and form a water chain from the bottom

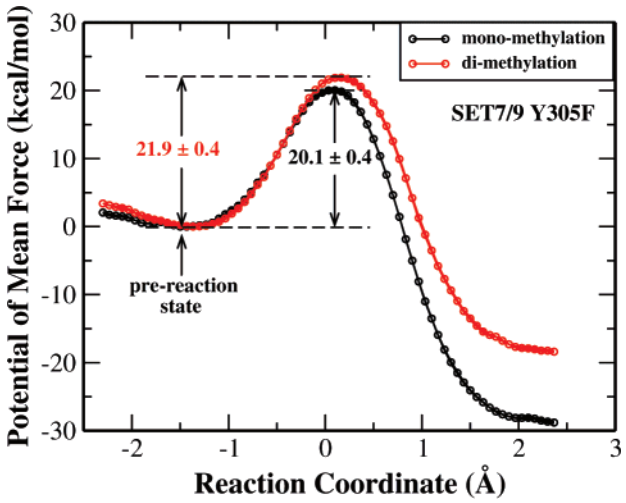


Figure 9. Potentials of mean force (PMF) for mono- and di-methylations catalyzed by SET7/9 Y305F mutant. The left side corresponds to the reactant and the right side is for the product. For each reaction, three PMF curves from different simulation time intervals (10−30 ps, 10−20 ps, and 20−30 ps) have been plotted.

of channel (lysine nitrogen) to the surface. The nearest water in the active site directly forms a hydrogen bond with the lone pair electron of substrate lysine nitrogen, which is very unfavorable for the methylation reaction. During the reaction process, the lone pair electron of substrate lysine nitrogen needs to be desolvated and its hydrogen bond network with water molecules needs to be reorganized, which leads to a higher activation barrier for the di-methylation reaction in SET7/9. For the case of the LSMT, it has a more spacious binding pocket, so that it can accommodate the methylated substrates without opening up its active site as well as its binding channel. Thus, LSMT has both mono- and di- methylation activities.

In order to further examine our insights into methylation state specificity, we have carried out ab initio QM/MM molecular dynamics simulations with umbrella sampling on both mono- and di-methylation catalyzed by the SET7/9 Y305F mutant. Tyr305 is known to be a key residue in controlling the methylation state specificity of SET7/9.[19] Our results in Figure 9 indicated that the calculated reaction barriers are 20.1 ± 0.4 kcal/mol and 21.9 ± 0.4 kcal/mol for the mono- and di-methylation catalyzed by the SET7/9 Y305F mutant, respectively. These results are consistent with the experimental results[19] that SET7/9 Y305F mutant not only has a high efficiency for mono-methylation, but also becomes a dimethylase.[19] Since the Y305F mutation leads to a less tight active site, we found that there is no water chain formed to allow the access of solvent water molecules to the enzyme active site for both methylation reactions catalyzed by the SET7/9 Y305F mutant. These results further support the proposed dynamic mechanism regarding how the methylation state specificity is controlled.

4. Conclusions

By employing ab initio QM/MM molecular dynamics simulations and an umbrella sampling method, we have investigated both mono- and di-methylations in two SET-domain protein lysine methyltransferases: SET7/9 and LSMT. Our simulation results confirm that SET7/9 is a mono-methylase, while LSMT is capable of catalyzing both mono- and di-methylation reactions, and indicates that the methyl-transfer reaction step is an important step in controlling the methylation state specificity. For the di-methylation reaction in SET7/9, it is found that the binding of the methylated lysine substrate in its active site opens up the cofactor AdoMet binding channel so that solvent water molecules get access to the active site. The consequence is that a water chain is formed from the enzyme active site to the protein surface, and in the reactant state, a solvent water molecule directly forms a hydrogen bond with the lone pair electron of substrate lysine nitrogen atom. Thus, comparing with either mono-methylation in SET7/9 or methylations in LSMT, extra energy is required to break the N···H−O hydrogen bond as well as to push a water molecule out of the channel during the di-methylation reaction in SET7/9, which results in a higher reaction activation barrier. As the di-methylation reaction in SET7/9 shares some similar features with lysine methylation reaction in aqueous solution, this indicates that the catalytic machinery of SET7/9 is impaired to some extent for the di-methylation reaction. Alternatively, for the LSMT, its active site is more spacious than that of SET7/9 so that the methylated lysine substrate can be accommodated without interfering with its catalytic power. In conclusion, our theoretical studies here provide new detailed insights regarding how the remarkable methylation state specificity is achieved in PKMTs. The results have taken account of the dynamic nature of enzyme system and are consistent with available experimental results as well as recent theoretical findings regarding the catalytic power of SET7/9.

Acknowledgment. This work has been supported by National Institute of Health (R01-GM079223). Y.Z. is also grateful for support from the National Science Foundation (CHE-CAREER-0448156, CHE-MRI-0420870), NYSTAR (James D. Watson Young Investigator Award), and NYU (Whitehead Fellowship for Junior Faculty in Biomedical and Biological Sciences). We thank NYU−ITS for providing computational resources and support.

Supporting Information Available: Complete refs 69 and 70. This material is available free of charge via the Internet at http://pubs.acs.org.

JA075896N