J|A|C|S
ARTICLES
Published on Web 01/06/2006

Catalytic Mechanism and Product Specificity of the Histone Lysine Methyltransferase SET7/9: An ab Initio QM/MM-FE Study with Multiple Initial Structures

Po Hu and Yingkai Zhang*

Contribution from the Department of Chemistry, New York University, New York, New York 10003

Received September 7, 2005; E-mail: yingkai.zhang@nyu.edu

Abstract: Histone lysine methylation is emerging as an important mechanism to regulate chromatin structure and gene activity. To provide theoretical understanding of its reaction mechanism and product specificity, ab initio quantum mechanical/molecular mechanical free energy (QM/MM-FE) calculations and molecular dynamics simulations have been carried out to investigate the histone lysine methyltransferase SET7/9. It is found that the methyl-transfer reaction catalyzed by SET7/9 is a typical in-line S_N2 nucleophilic substitution reaction with a transition state of 70% dissociative character. The calculated average free energy barrier at the MP2(6-31+G*) QM/MM level is 20.4 ± 1.1 kcal/mol, consistent with the activation barrier of 20.9 kcal/mol estimated from the experimental reaction rate. The barrier fluctuation has a strong correlation with the nucleophilic attack distance and angle in the reactant complex. The calculation results show that the product specificity of SET7/9 as a monomethyltransferase is achieved by disrupting the formation of near-attack conformations for the dimethylation reaction.

1. Introduction

In eukaryotic cells, DNA is tightly wrapped around an octamer of histone proteins and compacted into a dense structure known as chromatin. To access genetic information which is required in numerous cellular processes including DNA replication, gene expression, and DNA repair, chromatin needs to be partially unwound. One essential mechanism to regulate chromatin structure and thus to control the access of genomic DNA is methylation of lysine residues in the N-terminal histone tails.[1−3] Failure of appropriate methylation can lead to aberrant gene regulation and is implicated in a variety of human diseases, notably cancer.[4] The enzymes responsible for this important biological process are histone lysine methyltransferases (HKMTs), which catalyze the transfer of methyl group(s) from the cofactor S-adenosyl-methionine (AdoMet) to some specific lysine residues in the N-terminal histone tails.[5,6] With one exception of Dot1,[7] all known HKMTs contain the SET domain which represents a novel structural fold.[5] Although each lysine can be either mono-, di-, or trimethylated, a distinct histone lysine methyltransferase often only transfers a certain number of methyl group(s) to its target lysine residue, which is called product specificity.[5,6] Distinct methylation states have differing effects on chromatin structure and transcription.[3,5]

A fundamental question regarding the HKMT is its catalytic mechanism, i.e. how it catalyzes the transfer of methyl group(s) to a specific histone lysine residue. This methyl-transfer step is not only a key event of gene expression control, but also has high relevance to human disease. Despite extensive experimental studies,[8−17,20] some detailed mechanistic issues including the nature of the transition state have not yet been fully addressed. To our best knowledge, no theoretical study has been carried out to study histone lysine methylation. Here we focus on SET7/9, a monomethyltransferase which catalyzes the transfer of one methyl group to the unmodified histone lysine residue H3-K4,[8,11]

(1) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41−45.
(2) Lachner, M.; Jenuwein, T. *Curr. Opin. Cell Biol.* **2002**, *14*, 286−298.
(3) Sims, R. J.; Nishioka, K.; Reinberg, D. *Trends Genet.* **2003**, *19*, 629−639.
(4) Schneider, R.; Bannister, A. J.; Kouzarides, T. *Trends Biochem. Sci.* **2002**, *27*, 396−402.
(5) Xiao, B.; Wilson, J. R.; Gamblin, S. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 699−705.
(6) Cheng, X.; Collins, R. E. Zhang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267−294.
(7) Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. *Cell* **2003**, *112*, 711−723.
(8) 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.
(9) Trievel, R. C.; Beach, B. M.; Dirk, L. M.; Houtz, R. L.; Hurley, J. H. *Cell* **2002**, *111*, 91−103.
(10) Min, J.; Zhang, X.; Cheng, X.; Grewal, S. I.; Xu, R. *Nat. Struct. Biol.* **2002**, *9*, 828−832.
(11) 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.
(12) 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.
(13) Trievel, R. C.; Flynn, E. M.; Houtz, R. L.; Hurley, J. H. *Nat. Struct. Biol.* **2003**, *10*, 545−552.
(14) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12*, 177−185.
(15) 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.
(16) Couture, J. F.; Collazo, E.; Brunzelle, J. S.; Trievel, R. C. *Genes Dev.* **2005**, *19*, 1455−1465.
(17) Yin, Y.; Liu, C.; Tsai, S. N.; Zhou, B.; Ngai, S. M.; Zhu, G. *J. Biol. Chem.* **2005**, *280*, 30025−30031.
(18) Chuikov, S.; Kurash, J. K.; Wilson, J. R.; Xiao, B.; Justin, N.; Ivanov, G. S.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S. J.; Barlev, N. A.; Reinberg, D. *Nature* **2004**, *432*, 353−360.
(19) Kouskouti, A.; Scheer, E.; Staub, A.; Tora, L.; Talianidis, I. *Mol. Cell* **2004**, *14*, 175−182.
(20) Zhang, X.; Tamaru, H.; Khan, S. I.; Horton, J. R.; Keefe, L. J.; Selker, E. U.; Cheng, X. *Cell* **2002**, *111*, 117−127.

10.1021/ja056153+ CCC: $33.50

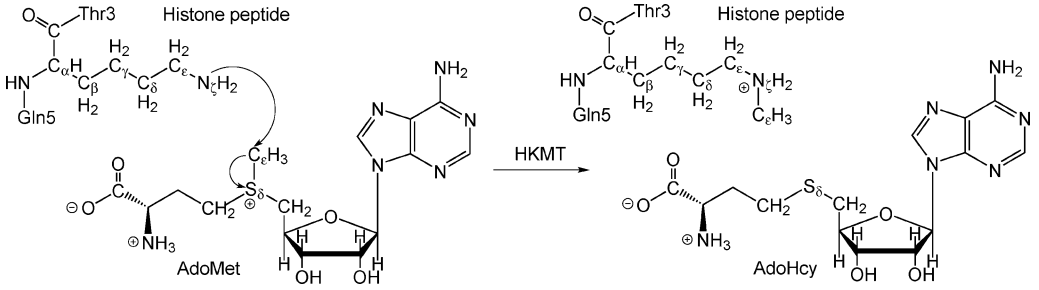

Figure 1. Methyl-transfer reaction catalyzed by SET7/9, in which one methyl group is transferred from AdoMet to the histone-lysine residue H3-K4.

as shown in Figure 1. Recently SET7/9 has also been found to methylate the transcription factor p53[18] and TAF10.[19] Like all other SET-domain HKMTs, the cofactor AdoMet and the substrate peptide bind to opposite faces of the SET7/9 and are connected by a narrow channel which has a hydrophobic inner wall.[11,12] The target lysine residue is inserted into this narrow channel to access the methyl moiety of AdoMet. Based on structural studies[8−17,20] that there is no apparent candidate for the catalytic base in the active site of SET-domain HKMTs, and biochemical studies[8,20] that SET7/9 and DIM-5 are active at pH 8 or higher and have an unusually high pH optimum of 10, the current general consensus is that the substrate lysine side chain is deprotonated when it binds to SET-domain-containing HKMT in the presence of AdoMet.[5,6] The active site of SET7/9 is dominated by several tyrosine residues, among which Tyr245 and Tyr305 have been identified as two key residues in controlling product specificity of SET7/9.[11,14] Tyr245 forms a hydrogen bond with substrate lysine, which is suggested to preorganize local conformation in the active site. Y245A mutation leads to a decreased activity with respect to unmodified lysine; however, it has substantial activity in dimethylation and trimethylation.[11]

In the present study, we have carried out multiple ab initio quantum mechanical/molecular mechanical free energy (QM/MM-FE) calculations and molecular dynamics simulations to investigate the methyl-transfer reaction catalyzed by SET7/9. Our calculations have characterized its reaction mechanism, elucidated the specific role of some key residues, and provided theoretical understanding of its product specificity.

2. Methods

Our theoretical approaches center on the combined ab initio quantum mechanical/molecular mechanical (QM/MM) method,[21−29] which allows for accurate modeling of the chemistry at an enzyme active site while properly including the effects of protein environment. The QM/MM interface was treated by a pseudobond approach.[30,31] To take account of enzyme dynamics, different snapshots of the enzyme−substrate complex from a molecular dynamics trajectory have been used as initial structures for QM/MM studies.[32−35] Enzyme reaction paths were determined by B3LYP/(6-31G*) QM/MM calculations with an iterative minimization procedure and the reaction coordinate driving method.[36] For determined reaction paths, single point MP2 QM/MM calculations with both 6-31G* and 6-31+G* basis sets have been carried out. Free energies along the reaction paths were determined by free energy perturbation calculations and the harmonic approximation.[24,36] The pseudobond ab initio QM/MM approach has been demonstrated to be powerful in the study of several enzymes, including enolase,[37] acetylcholinesterase,[32,38] 4-oxalocrotonate tautomerase,[39] and kinase.[40] Some theoretical predictions[37,39] were subsequently confirmed by experimental studies.[41−43]

The starting model of the enzyme−substrate complex was constructed on the basis of the crystal structure 1O9S,[11] which is a ternary complex of SET7/9 with S-adenosyl-homocysteine (AdoHcy) and a methylated histone peptide solved to 1.7 Å resolution. Although two almost identical chains are presented in the crystal structure 1O9S, we have only retained those atoms with chain label A and its corresponding substrates since HKMT SET7/9 is known as a monomer in solution. The initial structure of the reactant complex was constructed by moving the methyl group from methyl lysine residue to the sulfur atom of AdoHcy. The initial position of the methyl group was modeled on the basis of the geometry parameters of AdoMet bound to SET7/9 reported by Kwon et al.[12] We have also retained those crystal waters which are less than 4 Å away from the enzyme complex. Missing hydrogen atoms were added via InsightII.[44] The enzyme was solvated with a 27.0 Å solvent water sphere, centered on the active site (the sulfur atom of AdoMet). The prepared system has 9561 atoms, including 1841 water molecules. To take account of possible structural reorganizations and protein dynamics, the prepared system was first equilibrated with a series of minimizations interspersed by short molecular dynamics simulations, and then a 2.5 ns MD simulation was carried out. Eight snapshots were picked from 300 to 1000 ps at every 100 ps, and another three from 1.5, 2.0, and 2.5 ns. Each of 11 snapshots was first minimized by the MM method, and then optimized with B3LYP(6-31G*) QM/MM calculations using an iterative minimization approach,[36] leading to an optimized structure for the reactant.

In QM/MM calculations, the QM subsystem consists of AdoMet and the side chain of histone lysine residue H3-K4. With the $C_\alpha-C_\beta$

(21) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(22) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.
(23) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(24) Zhang, Y.; Liu, H.; Yang, W. In *Methods for Macromolecular Modeling*; Schlick, T., Gan, H. H., Eds.; Springer-Verlag: New York, 2002; pp 332−354.
(25) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(26) Cui, Q.; Karplus, M. *Adv. Protein Chem.* **2003**, *66*, 315−372.
(27) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.
(28) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.
(29) Zhang, Y. *Theor. Chem. Acc.* **2005**. In press.
(30) Zhang, Y.; Lee, T.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46−54.
(31) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
(32) Zhang, Y.; Kua, J.; McCammon, J. A. *J. Phys. Chem. B* **2003**, *107*, 4459−4463.
(33) Thorpe, I. F.; Brooks, C. L. *J. Am. Chem. Soc.* **2005**, *127*, 12997−13006.
(34) Klahn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645−15650.
(35) Claeyssens, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. *Chem. Commun.* **2005**; pp 5068−5070.
(36) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483−3492.
(37) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570.
(38) Zhang, Y.; Kua, J.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 10572−10577.
(39) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384−10393.
(40) Cheng, Y.; Zhang, Y.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 1553−1562.
(41) Poyner, R. R.; Larsen, T. M.; Wong, S. W.; Reed, G. H. *Arch. Biochem. Biophys.* **2002**, *401*, 155−163.
(42) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. *Biochemistry* **2004**, *43*, 6885−6892.
(43) Metanis, N.; Brik, A.; Dawson, P. E.; Keinan, E. *J. Am. Chem. Soc.* **2004**, *126*, 12726−12727.
(44) *InsightII*, Version 2000.1; Accelrys Software Inc.,: San Diego, CA, 2000.

bond of H3-K4 treated as a pseudobond,[30] the resulting QM subsystem has 66 atoms. All other atoms are described classically. Since this methyl-transfer reaction involves the breaking of the $S_\delta-C_\epsilon$ bond and the forming of the $C_\epsilon-N_\zeta$ bond as shown in Figure 1, we chose $R_{C\epsilon-N\zeta}-R_{S\delta-C\epsilon}$ as the reaction coordinate (RC). An iterative restrained optimization procedure[36] was then repeatedly applied to different points along the reaction coordinate, resulting in a minimum energy path. For obtained stationary points, Hessian matrixes for degrees of freedom involving atoms in the QM subsystem were calculated with B3LYP/6-31G* QM/MM calculations, and the corresponding vibrational frequencies were determined. The energy maximum on the path with one and only one imaginary frequency was located as the transition state, while reactant and product were characterized as having no imaginary frequency. The contribution of the QM subsystem fluctuation to the free energy change was then calculated with obtained frequencies using the harmonic approximation. In addition, for the determined reaction paths, single point MP2 calculations with both 6-31G* and 6-31+G* basis sets were carried out. Finally, the free energy change associated with QM/MM interaction were determined by the free energy perturbation (FEP) method. In FEP calculations, sampling of the MM subsystem was carried out with the QM subsystem frozen at different states along the reaction path.[36] The total free energy difference between the transition state and the reactant was obtained from the following formula:

$$\Delta F(R \rightarrow TS) = \Delta F_{qm/mm}(R \rightarrow TS) + \Delta E_{qm}(R \rightarrow TS) + \Delta F_{qm}^{fluctuation}(R \rightarrow TS) \quad (1)$$

where $\Delta F_{qm/mm}(R \rightarrow TS)$ is the free energy change associated with the QM/MM interaction, $\Delta E_{qm}(R \rightarrow TS)$ refers to the QM energy difference between two QM subsystems, and $\Delta F_{qm}^{fluctuation}(R \rightarrow TS)$ is the change in contribution from the QM subsystem fluctuation to the free energy difference.[24,36]

All calculations were carried out by modified versions of the Gaussian03[45] and TINKER programs.[46] Since our focus is on the active site, only those atoms within 20.0 Å of the sulfur atom of AdoMet were allowed to move. No cutoff for nonbonded interactions was used in the QM/MM calculations. Throughout the calculations, the pseudobond was treated with the 6-31G* basis set and its corresponding effective core potential parameters.[30] Empirical scale factor 0.9804[47] was used in frequency calculations. In all calculations, AMBER 95 all-atom force field[48,49] and the TIP3P model for water[50] were employed. Atomic charges on AdoMet in its zwitterion form have been determined with HF/6-31G* calculations and the restrained electrostatic potential (RESP) module in AMBER7.0.[48] In molecular dynamics and free energy perturbation calculations, the time step used was 2 fs, and bond lengths involving hydrogen atoms were constrained. A twin-range cutoff method was used to treat nonbonded interactions, with a long-range cutoff of 12 Å, a short-range cutoff of 8 Å, and the nonbonded pair list updated every 20 steps. In MD simulations, the temperature was maintained at 300 K using the weak coupling method with a coupling time of 0.1 *ps*. Each of FEP calculation consisted of 50 *ps* of equilibration and 300 *ps* of sampling. The final relative free energies were taken as the average of the "forward" and "backward" perturbation results.

In the simulation of mutated enzymes, point mutations to alanine were preformed by keeping the backbone unchanged and replacing the side chain by a methyl group. In preparation of starting geometries for the enzyme−MeLys complexes, one hydrogen atom of N_ζ of Lys4 was replaced with a methyl group, and the N−C bond distance was changed to the same as that in the crystal structure 1O9S.[11] Except for these modifications, no additional arrangements of the active site were made. The equilibration and simulation procedures were the same as that for the wild type.

3. Results and Discussions

Free Energy Reaction Barrier and Methylation Mechanism. For each of 11 initial structures from 2.5 ns MD simulations, we have calculated its reaction energy path for the transfer of methyl group from AdoMet to the substrate H3-K4 with B3LYP(6-31G*) QM/MM calculations, as shown in Figure 2. Both forward and backward reaction coordinate driving calculations yield consistent curves, and all reaction paths are smooth and continuous. Although different initial structures lead to different energy barriers, the locations of the transition states are very consistent. For each determined path, single point MP2 QM/MM calculations with both 6-31G* and 6-31+G* basis sets were carried out, and free energy differences were calculated. In comparison with calculated potential energy barriers as shown in Table 1, the average value of free energy barriers is about 1.1 kcal/mol lower, and their fluctuation is smaller. The calculated average free energy reaction barrier is 20.4 ± 1.4 kcal mol^{-1} and 20.7 ± 1.4 kcal mol^{-1} for MP2(6-31+G*)/MM and MP2(6-31G*)/MM calculations, respectively, which is in excellent agreement with the activation barrier of 20.9 kcal mol^{-1} estimated from the experimental value of k_{cat}[9] by the simple-transition state theory: $k(T) = (k_B T/h) \exp(-\Delta G^{\ddagger}/RT)$. The average barrier with B3LYP(6-31G*)/MM calculations is about 7 kcal mol^{-1} lower than the experimental value, which is consistent with the known fact that DFT methods often underestimate methyl-transfer reaction barriers.[51] From Table 1, it is obvious that a barrier from any single calculation is not necessarily close to the average. These results indicate that employing multiple starting structures in ab initio QM/MM-FE calculations is more robust than using a single starting structure.

QM/MM calculations from different starting structures yield a consistent mechanistic picture for this methyl-transfer reaction: **a typical in-line S_N2 nucleophilic substitution reaction with a mainly dissociative transition state**. From Table 2 and Figure 3, we can see that in the transition state, the CH_3 plane is almost at the middle of the attacking N_ζ atom and the leaving S_δ atom. The average distances of $S_\delta\cdots C_\epsilon$ and $C_\epsilon\cdots N_\zeta$ bonds are 2.32 ± 0.02 and 2.30 ± 0.02 Å, respectively, which are much longer than their covalent bond distances. Thus, this methyl-transfer transition state can be considered to be loose. For the $S_\delta\cdots C_\epsilon\cdots N_\zeta$ angle, it increases as the reaction progresses and has a maximal value of 173.3 ± 1.4° at the transition state. It has been proposed that a linear geometry of $S_\delta\cdots C_\epsilon\cdots N_\zeta$ is required for the methylation reaction.[52] Our calculation results indicate that an almost in-line conformation is indeed formed in the transition state, but not necessarily in the reactant state. The loose character of the transition state is further confirmed

(45) Frisch, M. J. et al. *Gaussian 03*, Revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.
(46) Ponder, J. W. *TINKER, Software Tools for Molecular Design*, version 3.6. The most updated version for the TINKER program can be obtained via the Internet at http://dasher.wustl.edu/tinker., maintained by J. W. Ponder. Accessed February 1998.
(47) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1996.
(48) 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.
(49) Markham, G. D.; Norrby, P. O.; Bock, C. W. *Biochemistry* **2002**, *41*, 7636−7646.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(51) Deng, L.; Branchadell, V.; Ziegler, T. *J. Am. Chem. Soc.* **1994**, *116*, 10645−10656.
(52) Coward, J. K. in *The Biochemistry of Adenosylmethionine*; Salvatore, F., Borek, E., Zappia, V., Williams-Ashman, H. G., Schlenk, F., Eds.; Columbia University Press: New York, 1977; pp 127−144.

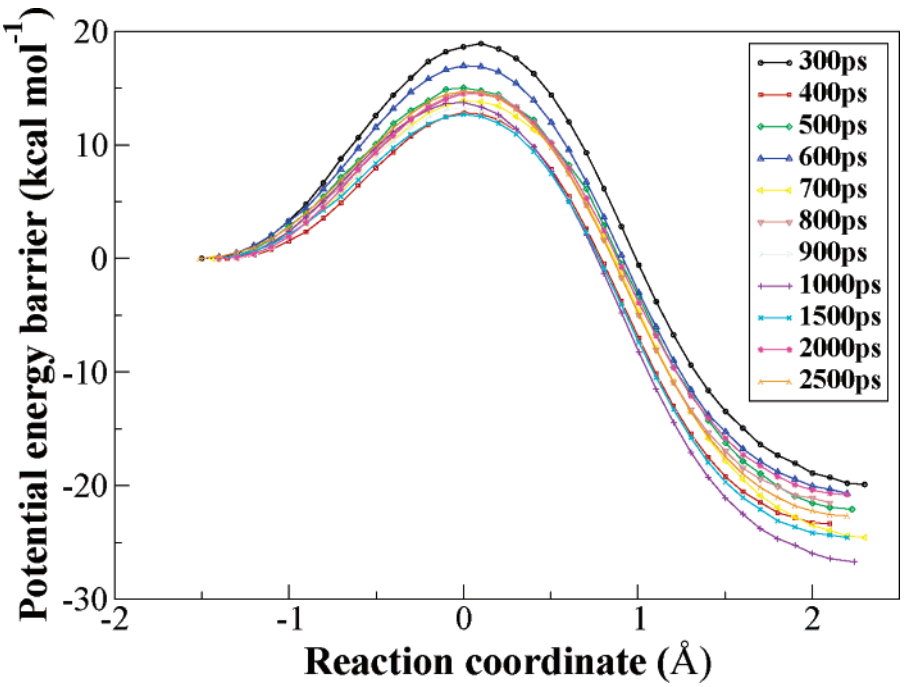

Figure 2. B3LYP(6-31G*) QM/MM minimum energy paths for the methyl-transfer step catalyzed by SET 7/9.

Table 1. Calculated Potential Energy Reaction Barriers (ΔE)[a] and Free Energy Reaction Barriers (ΔF)[a] for the Methyl-Transfer Reaction Catalyzed by SET7/9

	B3LYP(6-31G*)/MM		MP2(6-31G*)/MM		MP2(6-31+G*)/MM	
snapshots	ΔE	ΔF	ΔE	ΔF	ΔE	ΔF
300	18.6	15.0	26.4	22.7	26.0	22.3
400	12.8	12.1	20.0	19.3	19.7	19.0
500	15.0	13.6	22.0	20.6	21.7	20.2
600	17.0	15.6	24.7	23.5	24.3	22.9
700	13.9	13.0	21.4	20.6	21.2	20.3
800	14.7	12.7	21.6	19.6	21.1	19.1
900	13.3	12.3	19.9	19.0	19.4	18.5
1000	13.7	13.4	21.2	20.9	20.8	20.4
1500	12.7	13.3	19.8	20.5	19.3	20.0
2000	14.5	12.3	21.3	19.1	21.4	19.2
2500	14.7	14.8	22.1	22.2	22.0	22.1
average	14.6 ± 1.7	13.5 ± 1.1	21.9 ± 1.9	20.7 ± 1.4	21.5 ± 1.9	20.4 ± 1.4

[a] The unit is kcal mol^{-1}. The estimated activation barrier from the experimental value of k_{cat} is 20.9 kcal mol^{-1} using transition-state theory.

Table 2. Key Geometry Elements, Bond Lengths (Å) of d ($C_\epsilon \cdots N_\zeta$), $S_\delta \cdots C_\epsilon$, and Angle (deg) of θ ($S_\delta \cdots C_\epsilon \cdots N_\zeta$) in Both Reactant and Transition State

	reactant		transition state		
snapshots	d	θ	d	$S_\delta \cdots C_\epsilon$	θ
300	3.33	150.6	2.32	2.37	173.7
400	3.19	158.7	2.28	2.32	173.8
500	3.23	152.1	2.32	2.30	173.9
600	3.33	150.4	2.32	2.35	171.2
700	3.27	167.0	2.33	2.35	176.2
800	3.19	148.2	2.28	2.31	172.0
900	3.15	148.2	2.29	2.30	170.8
1000	3.24	150.7	2.35	2.30	173.0
1500	3.24	152.8	2.29	2.32	173.9
2000	3.19	153.3	2.27	2.31	173.9
2500	3.34	154.3	2.31	2.34	174.2
average	3.25 ± 0.06	153.3 ± 5.2	2.30 ± 0.02	2.32 ± 0.02	173.3 ± 1.4

by the bond order analysis.[53] As shown in Figure 4, the Wiberg bond order[54] diagram for this methyl-transfer reaction is always above the diagonal during the reaction process. It indicates that the breaking of the $S_\delta \cdots C_\epsilon$ bond is more advanced than the forming of the $C_\epsilon \cdots N_\zeta$ bond. At the transition state, the bond orders of $S_\delta \cdots C_\epsilon$ and $C_\epsilon \cdots N_\zeta$ bonds are 0.58 ± 0.01 and 0.30 ± 0.01, respectively. Since the fractional associativity of a mechanism is defined by the bond order of the attacking group,[55] the methyl-transfer reaction catalyzed by SET7/9 can be quantified as 30% associative and 70% dissociative based on the calculated Wiberg bond order of 0.30 ± 0.01 for $C_\epsilon \cdots N_\zeta$ at the transition state.

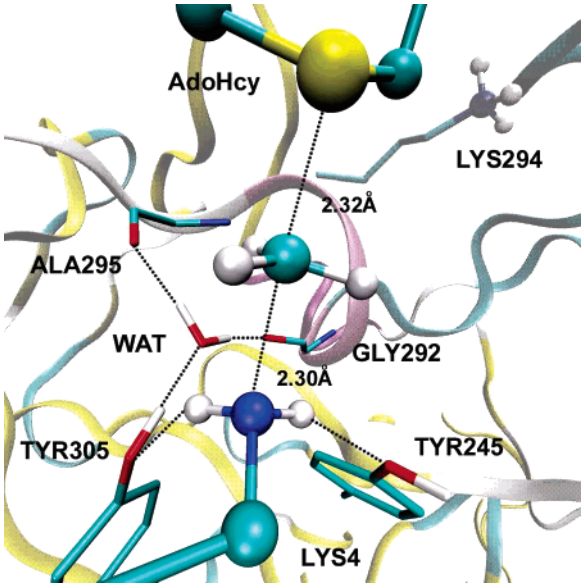

Figure 3. Structure of the HKMT SET7/9 active site at its transition state.

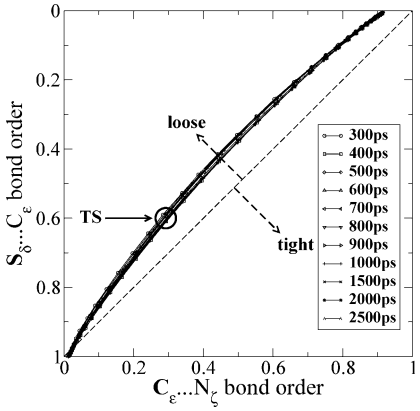

Figure 4. Wiberg bond order diagram for the methyl-transfer reaction catalyzed by HKMT SET7/9. The spots at the lower left corner represent the reactant state, whereas those at the upper right corner refer the product state. TS indicates the calculated transition state in the current study, which is above the diagonal.

To elucidate the structural origin of the barrier fluctuation, we have analyzed some key geometry elements in the reactants and the transition states, as shown in Table 2. It can be seen that the fluctuation of bond lengths and angles is much smaller in transition states than in reactants. Thus, we hypothesize that the free energy reaction barrier may be determined by the reactivity of the reactant conformation. Such a concept of reactive near-attack conformation has been employed in studies of catechol O-methyltransferase[56] and some other enzymes.[57] From the determined reaction mechanism, key geometry elements for such a methyl-transfer reaction are the $C_\epsilon\cdots N_\zeta$ distance d, and the $S_\delta\cdots C_\epsilon\cdots N_\zeta$ angle θ, which are the nucleophilic attack distance and angle, respectively. Indeed, we found that a combination of these two geometry elements, $d + 0.5(\cos(\theta))$, highly correlates with the calculated activation barrier, as shown in Figure 5. The correlation coefficients are

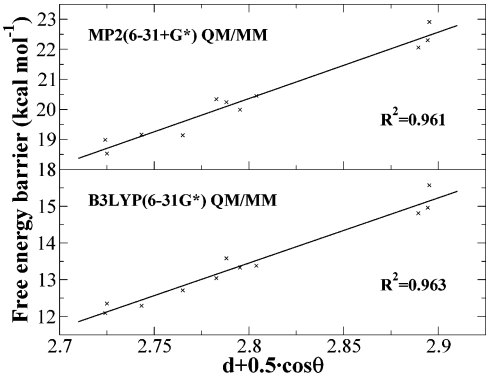

Figure 5. Calculated free energy barriers versus a combination of two key geometry elements: $C_\epsilon\cdots N_\zeta$ distance d, and $S_\delta\cdots C_\epsilon\cdots N_\zeta$ angle θ. B3LYP-(6-31G*) QM/MM results are presented in the lower panel (the correlation coefficient $R^2 = 0.963$), and MP2(6-31+G*) QM/MM results are shown in the upper panel (the correlation coefficient $R^2 = 0.961$).

0.963, 0.960, 0.961 for B3LYP(6-31G*), MP2(6-31G*), and MP2(6-31+G*) QM/MM-FE barriers, respectively. Therefore, $d + 0.5(\cos(\theta))$ can be used as an indicator for the reactivity of the reactant conformation. The smaller the $d + 0.5(\cos(\theta))$ is, the more reactive the reactant conformation will be. Meanwhile, we did not find such a strong correlation between the potential energy barrier and a combination of these two geometry elements. The correlation coefficient is merely 0.6 between the B3LYP(6-31G*) QM/MM potential energy barrier and $d + 0.5(\cos(\theta))$. Among all potential energy barriers, the one from the 300 ps snapshot is the highest which significantly deviates from the fitted curve. After a close examination, it is found that the hydrogen bond between Tyr305 and the crystal water, which is formed in all other 10 snapshots, is broken in the 300 ps snapshot. During free energy perturbation simulations, this hydrogen bond was reformed, and the free energy barrier is significantly reduced. These results demonstrate the advantage of FEP simulations which take into account the fluctuation of enzyme environment.

Origin of the Product Specificity. Unlike most other HKMTs, wild-type SET7/9 is a monomethyltransferase, which only transfers one methyl group from the AdoMet to the substrate H3-K4.[8,11] To provide theoretical understanding about why wild-type SET7/9 is short of dimethylation activity, a 2.5 ns MD simulation (WT-MeLys) has been carried out for wild-type SET7/9 in the presence of the methylated H3-K4 and AdoMet. In comparison with the corresponding MD simulation on SET7/9-Lys-AdoMet (WT-Lys), the distribution of $d + 0.5(\cos(\theta))$ in the WT-MeLys trajectory has been shifted to a much larger value, as shown in Figure 6. Thus, the complex conformation for wild-type SET7/9 in the presence of Me-Lys and AdoMet is unfavorable for the methyl-transfer reaction. These simulation results indicate that the product specificity of SET7/9 as a monomethylase is achieved by preventing the formation of near-attack conformations for the dimethylation reaction.

Since experimental studies have identified Tyr245 as a key residue in controlling product specificity of SET7/9, we have carried out MD simulation on Y245A-MeLys (Y245A mutant in the presence of the methylated H3-K4 and AdoMet). From Figure 6, it can be seen that in the Y245A-MeLys trajectory, the distribution of $d + 0.5(\cos(\theta))$ is very similar to that in the WT-Lys simulation, and near-attack reactive conformers are

(53) Takusagawa, F.; Fujioka, M.; Spies, A.; Schowen, R. L. In *Comprehensive Biological Catalysis: A Mechanistic Reference*; Sinnott, M., Ed.; Academic Press: New York, 1998; pp 1−30.
(54) Wiberg, K. B. *Tetrahedron* **1968**, *24*, 1083−1096.
(55) Mildvan, A. S. *Proteins* **1997**, *29*, 401−416.
(56) Lau, E. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 12387−12394.
(57) Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 139−148.

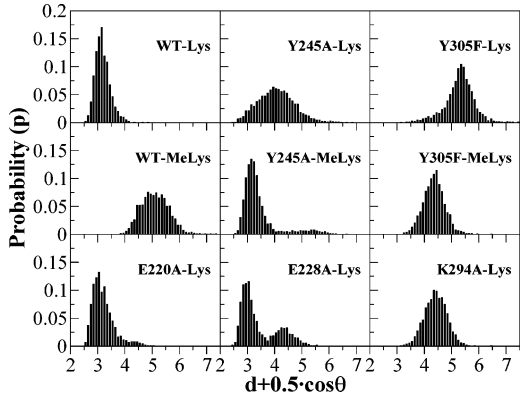

Figure 6. The distribution of $d + 0.5(\cos(\theta))$ calculated from MD simulation trajectories. d is the attack distance between AdoMet C_ϵ and lysine N_ζ, whereas θ is the attack angle $S_\delta \cdots C_\epsilon \cdots N_\zeta$. Snapshots are dumped every 1 ps from each 2.5 ns simulation after equilibration.

formed. This indicates that Y245A is able to catalyze methyl transfer from AdoMet to monomethylated lysine, which is consistent with experimental studies.[11] The analysis of three MD trajectories indicates that in the WT-MeLys simulation, Tyr245 precludes the rotation of methyl moiety of MeLys substrate, which causes some repulsions in the active site. As a result the substrate MeLys residue moves away from its previous position, which leads to a much longer $C_\epsilon \cdots N_\zeta$ distance. The mutation of Tyr245 to alanine relieves such steric effects; thus, there is enough room to accommodate the methyl group of the methyl-lysine. The attack distance $C_\epsilon \cdots N_\zeta$ becomes much shorter, and near-attack conformations for the methyl-transfer reaction are formed. These results are consistent with the early hypothesis that the product specificity of SET7/9 may be controlled by steric hindrance in the active site.[11,13,14] In addition, MD simulations on Y245A-Lys have also been carried out. We can see that the results in Figure 6 are very consistent with the experimental mutation results that Y245A mutant leads to a significant decreased activity with respect to the unmodified lysine while it has substantial activity for monomethylated lysine.[11]

Besides Tyr245, another key residue in controlling SET7/9's product specificity is Tyr305. Experimental studies have found that the Y305F mutant not only catalyzes the dimethylation reaction very efficiently but also has a higher catalytic activity for the monomethylation in comparison with the wild type.[14] With the similar procedure for Y245A systems, we have performed MD simulations on both Y305F-Lys and Y305F-MeLys. The simulation results indicate that the Y305F mutation leads to structural change in the active site, and near-attack conformations can seldom be formed in 2.5 ns MD trajectories for both Y305F-Lys and Y305F-MeLys systems as shown in Figure 6. Considering that the OH group of Tyr305 is critical to an important local hydrogen bonding network in the active site of the wild-type SET7/9 as we will discuss in the next section, it is not surprising that the Y305F mutation will disrupt this important hydrogen bond network and thus lead to structural change in the active site. Meanwhile, Cheng and his colleagues have found that the mutation of Phe281 to Tyr in DIM-5 can change the product specificity of DIM-5 from a trimethylase to a mono- and dimethylase.[14,58] In the enzyme active site, although two key residues Y245 and Y335 of SET7/9 superimpose well with two corresponding Tyr residues of DIM-5, it is not the case for Tyr305 of SET7/9 and Phe281 of DIM-5 (See Figure 4c in ref 14). Therefore, it may be reasonable to speculate that the mutation of Tyr305 to Phe may lead to the active site reorganization and thus result in the change of enzyme activity and product specificity. As the time scale of such large structural rearrangement is much longer than the current capability of MD simulations, to directly test this hypothesis requires further experimental structural studies on the Y305F mutant.

Table 3, Average MM Individual Residue Contributions to the Transition State Stabilization or Destabilization

residue	energy	residue	energy
Glu220	1.6 ± 0.2	Asp276	-1.3 ± 0.2
Glu228	3.6 ± 0.5	Glu279	1.9 ± 0.3
Tyr245	-1.3 ± 0.1	Lys294	-4.0 ± 1.0
Arg249	1.2 ± 0.0	Asp306	-1.5 ± 0.0
Asp256	-1.5 ± 0.1	Lys317	2.5 ± 0.1
Asn265	1.2 ± 0.7	Asp338	-1.2 ± 0.1
Thr266	-1.8 ± 0.8	Glu348	-1.2 ± 0.1
Asp270	-1.5 ± 0.1	Water	-3.0 ± 0.2

[a] All residues with $|\Delta E_i^{tot}| > 1.0$ kcal mol^{-1} are listed. They are calculated from all 11 TSs and reactants.

Roles of MM Residues in Enzyme Catalysis. To understand the role of individual residue in catalysis, van der Waals and electrostatic interactions between each MM residue and QM part have been calculated for both transition state and reactant, and their difference is the contribution of each MM residue to transition state stabilization. The formula is: $\Delta E_i = (E_{i/QM}^{vdw+ele})_{transition\ state} - (E_{i/QM}^{vdw+ele})_{reactant}$. Negative ΔE_i indicates that the residue i stabilizes the transition state, and vice versa. Since effects of conformational change and dielectric screening are absent in our analysis and original ab initio QM/MM calculations were performed variationally, we only use these numbers as qualitative indicators. It should be noted that a residue not stabilizing the transition state does not mean that it is not important for catalysis, since a residue can contribute to catalysis in many other ways, such as binding the substrate, facilitating the formation of near-attack conformations, and stabilizing the local or global structure which is necessary for catalysis. In our analysis, partial charges of QM atoms are obtained from B3LYP/6-31G* QM/MM calculations. (Charges from MP2 QM/MM calculations yield similar results.) All 11 TSs and reactants have been analyzed, and the results are very consistent, as presented in Table 3.

Among all residues, Lys294 contributes most to stabilize the transition state. Such a stabilizing effect is mainly due to the fact that the positive charge, which transfers from the AdoMet to the substrate lysine during the methyl-transfer reaction, moves away from the Lys294. This result is consistent with the mutation result that the K294A mutant significantly reduces its catalytic activity.[12] Lys294 belongs to the R/HxxNHS motif, which is the most conserved sequence found in members of the SET-domain-containing protein family.[12] Since Lys294 is found to be in close contact with AdoMet in crystal structures, it has been suggested that Lys294 plays an important role in constructing the AdoMet binding pocket and properly orienting AdoMet for the methyl-transfer reaction.[8,12] Our MD simulation on the K294A mutant does support this hypothesis. As shown in Figure 6, the distribution of $d + 0.5(\cos(\theta))$ for the K294A

(58) 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.

mutant is much broader than that for the wild-type enzyme, and the peak has been shifted to 4.5 Å. Therefore, our calculation results indicate that Lys294 may have dual roles in SET7/9 catalysis: one is to strongly stabilize the transition state by electrostatic interaction, and the other is to properly orient AdoMet for the methyl-transfer reaction. It should be noted that a basic residue corresponding to Lys294 in SET7/9 has not been found in the catalytic domain of a viral SET enzyme (vSET).[59] This raises an interesting question regarding how vSET catalyzes the methylation of Lys27 in histone H3, which warrants further experimental and theoretical studies.

Besides amino acid residues, one crystal water in the active site also strongly stabilizes the transition state. In early structural studies,[11] this crystal water has been hypothesized to play a critical role in catalysis by directly forming a hydrogen bond with the histone-lysine residue, which preorients the lysine N_ζ atom in a favorable conformation for nucleophilic attack. However, in our WT-Lys simulation, we found that this crystal water does not directly form a hydrogen bond with the lysine substrate, with the average $N_\zeta \cdots O$ distance of 3.82 ± 0.36 Å. Although this result does not support the hypothesis proposed in ref 11, it is actually quite consistent with the experimental structural data of 1O9S[11] in which the distance between the crystal water and N_ζ of the methylated lysine is 3.4 Å. Meanwhile, in most MD snapshots, this water molecule directly forms three hydrogen bonds with the hydroxyl group of Tyr305 and carbonyl oxygens of Gly292 and Ala295, respectively, which is the same as that in the crystal structure.[11] This stable local hydrogen bond network aligns the dipole of water along the direction favoring the methyl-transfer reaction, as shown in Figure 3.

From Table 3, it can be seen that Glu228 most strongly destabilizes the transition state. In crystal structures,[11,12] except for the formation of the salt bridge with Lys294, the side chain of Glu228 is exposed to the solvent and does not form any significant contact with other residues or substrates. Considering the importance of Lys294 as we discussed above, the role of Glu228 can be hypothesized to position Lys294 thus to maintain the local structure of the active site. However, our MD simulation on wild-type SET7/9 indicates that the salt bridge between Glu228 and Lys294 is very dynamic, and it is not always formed. Meanwhile, as shown in Figure 6, a significant number of near-attack conformations can still be formed during the simulation of the E228A mutant. It should be noted that the overall enzyme activity is not only dependent on the formation of near-attack conformations but also determined by the reaction barrier and other effects. Considering that Glu228 strongly destabilizes the transition state in the wild-type enzyme, Glu228 could be a target residue for mutation to improve the efficiency of SET7/9 by lowering the enzyme reaction barrier. Among other destabilizing residues in Table 3, Glu220 would also be a good candidate for such redesign purpose since its side chain is exposed to the solvent and it does not form any close contact with other residues or substrates. From Figure 6, we can also see that E220A mutant still has a high probability for the formation of near-attack conformers. It should be noted that these mutation suggestions are just theoretical hypotheses which require further experimental tests.

(59) Manzur, K. L.; Farooq, A.; Zeng, L.; Plotnikova, O.; Koch, A. W.; Zhou, M. M. *Nat. Struct. Biol.* **2003**, *10*, 187−196.

4. Conclusions

The methyl-transfer reaction catalyzed by histone lysine methyltransferase SET7/9 has been studied by an ab initio QM/MM-FE approach. Multiple snapshots of the enzyme−substrate complex from a molecular dynamics trajectory have been used as initial structures to determine reaction paths and free energy barriers. Our calculations have characterized its reaction mechanism: it is a typical in-line S_N2 nucleophilic substitution reaction with a transition state of 70% dissociative character. The calculated free energy barrier with MP2(6-31+G*) QM/MM calculations is 20.4 ± 1.1 kcal mol^{-1}, which is in excellent agreement with the activation barrier of 20.9 kcal mol^{-1} estimated from the experimental value of k_{cat}. It is found that the height of the free energy barrier is highly correlated to a combination of the nucleophilic attack distance $C_\epsilon \cdots N_\zeta$ (d) and the nucleophilic attack angle $S_\delta \cdots C_\epsilon \cdots N_\zeta$ (θ) in the reactant complex.

Although a lysine residue can be either mono-, di-, or trimethylated, wild-type SET7/9 is a monomethyltransferase which transfers only one methyl group to the unmodified histone-lysine residue H3-K4. Such remarkable product specificity is a hallmark of histone-lysine methylation, and is crucial for the histone code concept. Our calculations indicate that the product specificity of SET7/9 is achieved by disrupting the formation of near-attack conformations for the di-methylation reaction. From the QM/MM interaction energy decomposition analysis, Lys294 and a crystal water in the active site are found to greatly stabilize the transition state through electrostatic interaction. In addition, we suggest that Glu220 and Glu228 could be two promising targets for mutation in order to improve the catalytic efficiency of SET7/9.

Acknowledgment. Y.Z. is grateful for the support from National Science Foundation (CHE-CAREER-0448156, CHE-MRI-0420870) and NYSTAR (James D. Watson Young Investigator Award).

Supporting Information Available: Complete ref 45.

JA056153+