JCTC Journal of Chemical Theory and Computation

Understanding Energetic Origins of Product Specificity of SET8 from QM/MM Free Energy Simulations: What Causes the Stop of Methyl Addition during Histone Lysine Methylation?

Yuzhuo Chu,[†] Qin Xu,[†] and Hong Guo*,[†,‡]

Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996 and UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6164

Received December 2, 2009

Abstract: Biological consequences of histone lysine methylation depend on the methylation states of the lysine residues on the tails of histone proteins that are methylated by protein lysine methyltransferases (PKMTs). Therefore, the ability of PKMTs to direct specific degrees of methylation (i.e., product specificity) is an important property for regulation of chromatin structure and gene expression. Here, the free energy simulations based on quantum mechanical/molecular mechanical (QM/MM) potentials are performed for the first, second, and third methyl transfers from S-adenosyl-L-methionine to the ε-amino group of the target lysine/methyl lysine in SET8, one of the important PKMTs. The key questions addressed in this paper include the energetic origin of the product specificity and the reasons for the change of the product specificity as a result of the replacement of Tyr334 by Phe. The free energy barriers for the three methyl transfers in SET8 as well as in the mutant obtained from the simulations are found to be well correlated with the experimental observations on the product specificity of SET8 and the change of product specificity as a result of the mutation. The results support the suggestion that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. Furthermore, the stability of a water molecule to be located at the active site is examined under different conditions using the free energy simulations, and its role in controlling the product specificity is discussed. The QM/MM molecular dynamics (MD) simulations are also performed on the reactant complexes of the first, second, and third methyl transfers. The results show that the information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful indicators in the prediction of product specificity for PKMTs.

Introduction

The tails of core histone proteins in the nucleosome are subject to a variety of post-translational covalent modifications, and these modifications can be read by other proteins to lead to distinct downstream events in the regulation of chromatin structure and gene expression.[1] Protein lysine methyltransferases (PKMTs) catalyze one such modification, i.e., histone lysine methylation. Histone lysine methylation can govern a number of important biological processes, including heterochromatin formation, X-chromosome inactivation, and transcriptional silencing and activation.[2] Several lysine residues on histone proteins have been identified to be the sites of methylation, including histone H3 Lysine 4 (H3−K4), H3−K9, H3−K27, H3−K36, H3−K79, and H4−K20. In addition to selecting different lysine sites for methylation (i.e., substrate specificity), PKMTs may also

* Corresponding author e-mail: hguo1@utk.edu.
[†] University of Tennessee.
[‡] Oak Ridge National Laboratory.

10.1021/ct9006458
Published on Web 03/17/2010

differ in their ability to transfer one, two, or three methyl groups from S-adenosyl-L-methionine (AdoMet, the methyl donor) to the ε-amino group of the target lysine.[3] This property of the enzymes is called product specificity. Since biological consequences of histone lysine methylation depend on the number of methyl groups added to the lysine residues, understanding the energetic origins of product specificity and developing suitable strategies for manipulation of the signaling properties are of considerable interest.

Biochemical studies and structural analyses of the SET-domain PKMT complexes have identified a tyrosine/phenylalanine switch that can control the product specificity. Comparison of the active site structure of DIM-5 (a trimethylase)[4] with that of SET7/9[5] or SET8[6] (a monomethylase) showed that a single amino acid residue occupies a structurally similar position in the enzymes (e.g., F281 in DIM-5 and Y334 in SET8) and is in proximity of the ε-amino group of the target lysine.[4b,7] It has been demonstrated that DIM-5 can be converted from a K-9 trimethyltransferase to a K-9 mono/dimethyltransferase by the F281→Y mutation[4b] and that the SET7/9 Y305F[4b] or SET8 Y334F mutant[5a,7] was able to generate a dimethylated instead of a monomethylated lysine product. In each case, the substrate specificity was not changed, and the mutation had little effect on the overall reaction rate. Understanding the role of the residue at this tyrosine/phenylalanine switch position and the energetic origin for the change of the product specificity as a result of the mutation can provide important insights into the property of the PKMT product specificity. Systematic determinations of the structures for SET8 and its Y334F mutant complexed with an unmodified, monomethylated, or dimethylated H4−K20 peptide along with AdoHcy were undertaken[7] to pinpoint the structural origin for the existence of the tyrosine/phenylalanine switch. It was proposed on the basis of the existence and absence of an active-site water molecule in these structures along with some other information that the Phe/Tyr switch may regulate product specificity through altering the affinity of the observed water molecule and that the dissociation of this water molecule is likely to be essential for the multiple methylation process to proceed.[7] Nevertheless, the energetic interpretations for the product specificity of SET8 and its alternation due to the Y334→F mutation are still lacking, and it is not clear as to how the enzyme's ability to stabilize/destabilize the water molecule at the active site would change as a result of the mutation and/or methyl addition.

Computer simulations can provide important insights into the energetic origins of the product specificity as well as the effects of mutations. One approach that has been used previously for PKMTs is to perform the free energy (potential of mean force) simulations with hybrid quantum mechanical/molecular mechanical (QM/MM) potentials and to establish the correlations between the free energy profiles of the methyl transfers and the product specificity[8] (see below for the discussions of the results based on other computational approaches). In our earlier communication,[8b] it was demonstrated that, for DIM5, SET7/9, and their certain mutants, the three free energy barriers for the methyl transfers may be used in each case to explain the product specificity observed experimentally. It was hypothesized[8b] that the relative efficiencies of the chemical steps involving the three methyl transfers from AdoMet to the ε-amino group of the target lysine in PKMTs may determine, at least in part, the product specificity. The results of the QM/MM molecular dynamics (MD) simulations on the reactant complexes have also been compared for the first, second, and third methyl transfers for SET7/9, DIM-5, and their mutants.[8b] It was shown that a correlation may be established between the formation of the reactive configurations for the three methyl transfers and the product specificities of the enzymes. One problem in the earlier work[8b] is that the experimental structures for the SET7/9 and DIM-5 complexes with *different methylation states* for the target lysine residues do not exist, and some manual modifications had to be introduced in the generation of suitable reactant complexes containing methylated lysine. In order to establish the prediction on the relationship between the efficiency of the methyl transfers and product specificity, additional simulations need to be performed on the PKMT (and the mutant) complexes for which the X-ray structures have been determined at different methylation states for the target lysine residue. SET8 is an excellent system for such investigations because of the recent availability of several experimental structures with unmodified, mono- and dimethylated lysine residue at the active site. Furthermore, the location of the important active-site water molecule has also been clearly identified in the X-ray structures. The simulations based on these structures may not only lead to a better understanding of the energetic origin of the product specificity but also provide important energetic information concerning the stability of this water molecule at the active site at different stages of methylation in different systems that is believed to be a key property of the enzymes in controlling the product specificity.[7]

Here, we report the results of QM/MM free energy simulations on SET8 and its Y334F mutant. The free energy barriers for the methyl transfers in SET8 and the mutant obtained from the simulations are found to be well correlated with the experimental observations on the product specificities, supporting the suggestion that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. Furthermore, the stability of the water molecule at the active site under different conditions (see above) is also examined on the basis of the free energy simulations. The free energy profiles show that the stability of the water molecule at the active site decreases significantly as a result of the Y334→F mutation as well as the methyl addition to the lysine residue. Such changes are likely to make it easier for the water molecule to dissociate from the active site and create the space for further methyl addition. The QM/MM MD simulations are also performed on the reactant complexes of the first, second, and third methyl transfers. The results show that the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful

indicators in the prediction of product specificity for PKMTs, although other factors can be involved as well.

Methods

QM/MM free energy (potential of mean force) and MD simulations were applied to determine free energy profiles for the first, second, and third methyl transfers from AdoMet to the ε-amino group of the target lysine (methyl lysine) and to characterize the active-site dynamics of the reactant complexes of the methyl transfers in SET8 and the Y334F mutant using the CHARMM program.[9] AdoMet/AdoHcy and lysine/methyl-lysine side chains were treated by QM and the rest of the system by MM. The link-atom approach[10] as implemented in CHARMM was applied to separate the QM and MM regions. Although the QM/MM approach in principle is not required for MD investigations of the reactant complexes, the previous studies on SET7/9 and DIM-5[8] showed that the QM/MM MD approach seems to provide a good description of the active-site dynamic features of the reactant complexes that are consistent with experimental observations concerning the product specificity. A modified TIP3P water model[11] was employed for the solvent. The stochastic boundary molecular dynamics method[12] was used for the QM/MM MD and free energy simulations. The system was separated into a reaction zone and a reservoir region, and the reaction zone was further divided into a reaction region and a buffer region. The reaction region was a sphere with radius r of 20 Å, and the buffer region extended over 20 Å $\leq r \leq$ 22 Å. The reference center for partitioning the system was chosen to be the C_δ atom of the target lysine residue/methyl lysine. The resulting systems contained around 5500 atoms, including about 800−900 water molecules.

The SCC-DFTB method[13] implemented in CHARMM was used for the QM atoms, and the all-hydrogen CHARMM potential function (PARAM27)[14] was used for the MM atoms. High-level ab initio methods (e.g., B3LYP and MP2) are too time-consuming to be used for MD and free energy simulations. The results of the SCC-DFTB and B3LYP/6-31G** methods for the description of the methyl transfer in a small model system have been compared in earlier studies[8] using an energy-minimization-based approach. This comparison allowed us to understand the performance of the semiempirical method in the description of the bond breaking and making for the system under investigation. It was found that, although the SCC-DFTB optimized geometries along the reaction pathway seemed to be rather close to those from B3LYP/6-31G**, there are some systematic deviations of the SCC-DFTB method in the description of the energetics of the methyl transfer. To correct the errors due to the deficiency of the SCC-DFTB method, the empirical correction introduced in the earlier studies[8] was applied to the free energy curves obtained from the potential of mean force simulations in the present work (see below). In the earlier study of DIM-5 (see the Supporting Information in ref 8b), it was shown that that the energy curves from the corrected SCC-DFTB and B3LYP/6-31G** were very close, supporting the use of the approach with the empirical correction. It should be pointed out that the reason that the simple empirical correction can be used in this and previous studies is because the bond breaking and making events in this and the previous papers all involve simple and similar S_N2 methyl transfer processes so that most of the errors are expected to be canceled out. Moreover, the relative free energy barriers, as opposed to the absolute barriers, are expected to be more important in the determination of the product specificity. The relative free energy barriers are expected to be less sensitive to the choice of the QM method due to the cancellation of the errors. Indeed, the results of our earlier simulations[8a] were confirmed by the use of a quite different QM/MM approach[15] with a difference of only about 1 kcal/mol in the relative free energy barriers of the first and second methyl transfers in SET7/9.

The initial coordinates for the reactant complexes of the first, second, and third methyl transfers were based on the crystallographic complexes (PDB codes: 1ZKK, 3F9W, 3F9X, and 3F9Y) of SET8 and its Y334F mutant containing AdoHcy and short H4K20, H4K20me1, and H4K20me2 peptides.[5a,7] In all the cases, a methyl group was manually added to AdoHcy to form AdoMet. In addition, for the reactant complex of the second methyl transfer in the wild type, a methyl group was manually added to the target lysine in the X-ray structure (1ZKK) to form the methylated lysine. The initial structures for the entire stochastic boundary systems were optimized using the steepest descent (SD) and adopted-basis Newton−Raphson (ABNR) methods. The systems were gradually heated from 50.0 to 310.15 K over 50 ps. A 1 fs time step was used for integration of the equation of motion, and the coordinates were saved every 50 fs for analyses. The 1.5 ns QM/MM MD simulations were carried out for each of the reactant complexes of the first, second, and third methyl transfers, and the data from the final 0.5 ns were used to generate the distribution maps of $r(C_M-N_\xi)$ and θ in each case (see below). As discussed in the previous studies and mentioned earlier in this paper, the S_N2 methyl transfer from AdoMet to H4−K20, H4−K20me1, or H4−K20me2 is presumably more efficient if the S−CH_3 group of AdoMet is well aligned with the lone pair of electrons on N_ξ in the reactant complex, i.e., with a small θ angle and relatively short C_M-N_ξ distance.[8] Here, θ is defined as the angle between the direction of the C_M-S_δ bond (r_2) and the direction of the electron lone pair (r_1) (see Figure 1). Therefore, we determined the distributions of $r(C_M-N_\xi)$ and θ from the QM/MM MD trajectories to obtain the information about the relationship between these distributions and product specificity. Moreover, the histogram method was used to calculate the probability density distributions of $r(C_M-N_\xi)$ and θ and the relative free energies as functions of $r(C_M-N_\xi)$ and θ. For $r(C_M-N_\xi)$, histograms with a bin width of 0.1 Å were used, and the probability density in the ith histogram is as follows: $\rho_i = N_i/N$ (N is the total number of the configurations from the MD simulations, and N_i is the occurrence number in the ith histogram). To calculate the probability density distribution of θ, the histograms with a width of 10° were used, and N_i was weighted by $1/A_i$, where A_i is the area of the ith histogram of θ. Thus, the probability density in the ith histogram of θ is as follows: $N_i/(N \times A_i)$. The relative free energy of the ith histogram of $r(C_M-N_\xi)$ and θ were calculated through $W_i =$

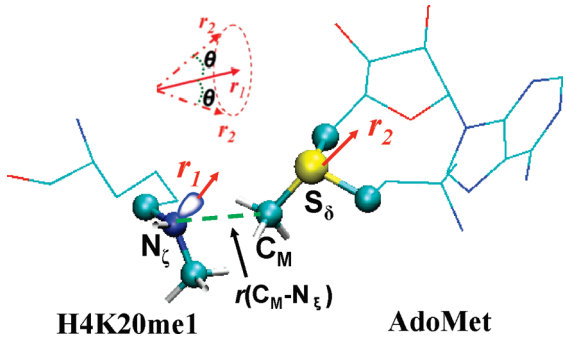

Figure 1. Definition of the structural parameters for monitoring the relative orientation of AdoMet and H4K20me1 [H4K20 and H4K20(me)$_2$] in the reactant complex. The efficiency of the methyl transfer may be related to the distributions of $r(C_M\cdots N_\zeta)$ and θ in the reactant complexes. θ is defined as the angle between the two vectors r_1 and r_2. Here, r_1 is the direction of the lone pair of electrons on N_ζ and r_2 is the vector pointing from C_M to S_δ. The reaction coordinate for calculating the free energy profiles for the methyl transfers is $R = r(C_M\cdots S_\delta) - r(C_M\cdots N_\zeta)$.

$-k_BT \times \ln \rho_i$, where k_B is Boltzmann's constant and T is the temperature (see the Supporting Information of ref 8a for additional information).

The umbrella sampling method[16] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM)[17] was applied to determine the change of the free energy (potential of mean force) as a function of the reaction coordinate for the methyl transfer from AdoMet to H4−K20, H4−K20me1, or H4−K9me2 in the wild-type and mutated enzymes. The reaction coordinate was defined as a linear combination of $r(C_M-N_\xi)$ and $r(C_M-S_\delta)$ [$R = r(C_M-S_\delta) - r(C_M-N_\xi)$] (see Figure 1). For each methyl transfer process, 20 windows were used, and for each window, 50 ps production runs were performed after 20 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 50 to 500 kcal mol^{-1} Å^{-2}. The statistical errors for the free energy profiles were also estimated and were found to be quite small (see the Supporting Information). In addition, the umbrella sampling method was applied to determine the free energy profiles for the movement of the active-site water molecule (W1) in the wild type and Y334F at different methylation states. The reaction coordinate was defined as the distance between the water oxygen and sulfur atom of AdoMet [i.e., $r(O_{w1}\cdots S_\delta)$]. Twenty windows were used for the change of the water position, and for each window, 20 ps production runs were performed after a 20 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 20 to 50 kcal mol^{-1} Å^{-2}. The distance for which the simulations were performed is in the range of 3−15 Å. The simulations were not performed for longer distances, because the use of the stochastic boundary method in the present study may limit the flexibility of the water molecule as it moves closer to the boundaries of the models (at ∼22−25 Å from sulfur atom). Additional simulations with different boundary conditions will be performed in the future.

Results

The average active-site structure of the reactant complex for the first methyl transfer in SET8 is given in Figure 2A. Figure 2A shows that the active-site structure has the lone pair of electrons on N_ζ of the target lysine well aligned with the methyl group of AdoMet. This is further demonstrated by the large population of the structures with relatively short $r(C_M\cdots N_\zeta)$ distances and small values of the θ angle as well as the free energy plots generated from the results of the simulations (Figure 2B). The average distance between N_ξ and the methyl group (C_MH_3) is approximately 3.0 Å, and the angle is mainly in the range of 0−30°. Figure 2A also shows that Tyr245 forms a hydrogen bond with the ε-amino group of the target lysine, and this hydrogen bond may help to orientate the direction of the electron lone pair toward the methyl group of AdoMet. A water molecule (W1) forms stable hydrogen bonds with the both ε-amino groups of H4K20 and Tyr334 (the important tyrosine/phenylalanine switch residue, see above). Figure 2D and E show that, for the reactant complex of the second methyl transfer, the average distance between N_ξ and the methyl group (∼4.5 Å) and the values of θ (mainly in the range of 45−120°) become significantly larger compared to those for the first methyl transfer (Figure 2A and B). Thus, the S−CH_3 group of AdoMet cannot be well aligned with the lone pair of electrons on N_ξ for the second methyl transfer, suggesting that the efficiency of the corresponding methyl transfer may be significantly compromised. Indeed, Figure 2E shows that the free energy cost for producing a structure like the one in Figure 2A is approximately 4−5 kcal/mol [i.e., ∼3 kcal/mol for changing $r(C_M\cdots N_\zeta)$ from 4.5 Å to 3 Å and 1.5 kcal/mol for changing θ to less than 30°].

The average structures for the first, second, and third methyl transfers in Y334F are given in Figure 3. As is evident from Figure 3C, the lone pair of electrons on N_ζ of the methyl lysine is well aligned with the methyl group of AdoMet for the second methyl transfer in Y334F. This is in contrast to the case for the second methyl transfer in the wild type for which the two cannot be well aligned (see above). The results suggest that the efficiency of the second methyl transfer is likely to be significantly enhanced due to the improvement of the reactant structure for the methyl transfer, although other factors may be involved as well (see below). For the third methyl transfer in Y334F (Figure 3E), the S−CH_3 group of AdoMet cannot be well aligned with the lone pair of electrons on N_ξ in the reactant complex as indicated by the long $r(C_M\cdots N_\zeta)$ distance (∼4.6 Å) and large values of the θ angle (45−150°). Thus, the corresponding methyl transfer is unlikely to be efficient. One of the key structural changes at the active site is the relocation of the active-site water molecule (W1). Indeed, W1 occupies a position in Figure 3C that is completely different from the one in Figure 2D (i.e., the complex for the second methyl transfer in the wild type) and is not in close contact with the target lysine/methyl lysine and the Y334/F334 residue anymore. For the third

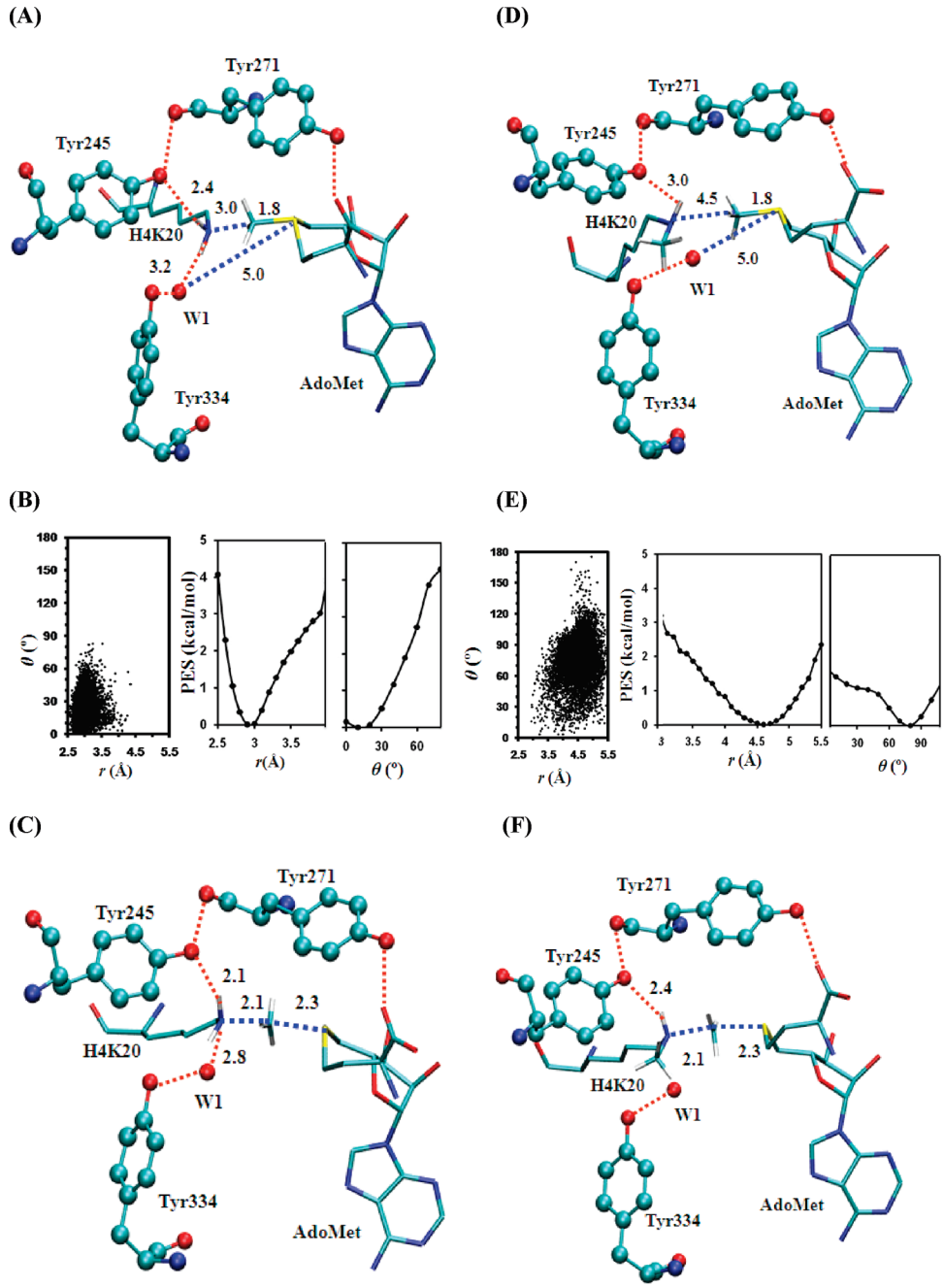

Figure 2. MD results for the wild-type enzyme (SET8). (A) The average active-site structure of the reactant complex for the first methyl transfer. SET8 is shown in balls and sticks, and AdoMet and the H4K20 side chain are in sticks. Hydrogen atoms are not shown for clarity, except for those on N_ζ and the transferable methyl group. Hydrogen bonds are indicated by red dotted lines, and the distances related to the reaction coordinates are also shown. (B) Left: the two-dimensional plot of $r(C_M \cdots N_\zeta)$ and θ distributions based on the 1.5 ns simulations of the reactant complex for the first methyl transfer. Middle: the free-energy change as a function of $r(C_M \cdots N_\zeta)$ obtained from the distributions. Right: the free-energy change as a function of θ obtained from the distributions. (C) The average structure near the transition state for the first methyl transfer obtained from the free energy (potential of mean force) simulations. (D) The average structure of the reactant complex for the second methyl transfer. (E) Left: the two-dimensional plot of $r(C_M \cdots N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer. Middle: the free-energy change as a function of $r(C_M \cdots N_\zeta)$ obtained from the distributions. Right: the free-energy change as a function of θ obtained from the distributions. (F) The average structure near the transition state for the second methyl transfer.

methyl transfer in Y334F, W1 has been pushed away from the active site during the MD simulations and is not visible in Figure 3E.

The free-energy profiles for the first and second methyl transfers in SET8 are plotted in Figure 4A as a function of the reaction coordinate; the free energy barrier for the first methyl transfer was calculated to be 15.8 kcal/mol. This free energy barrier is within the error limit of the average barrier from earlier single-point MP2/6-31G+G(d,p)/MM calculations (13.9 ± 2.3 kcal/mol) on SET8.[18] It should be pointed out, however, that cautions must be exercised when the energetic data obtained on the basis of very different computational approaches are compared. As is evident from Figure 4A, the free energy barrier for the second methyl

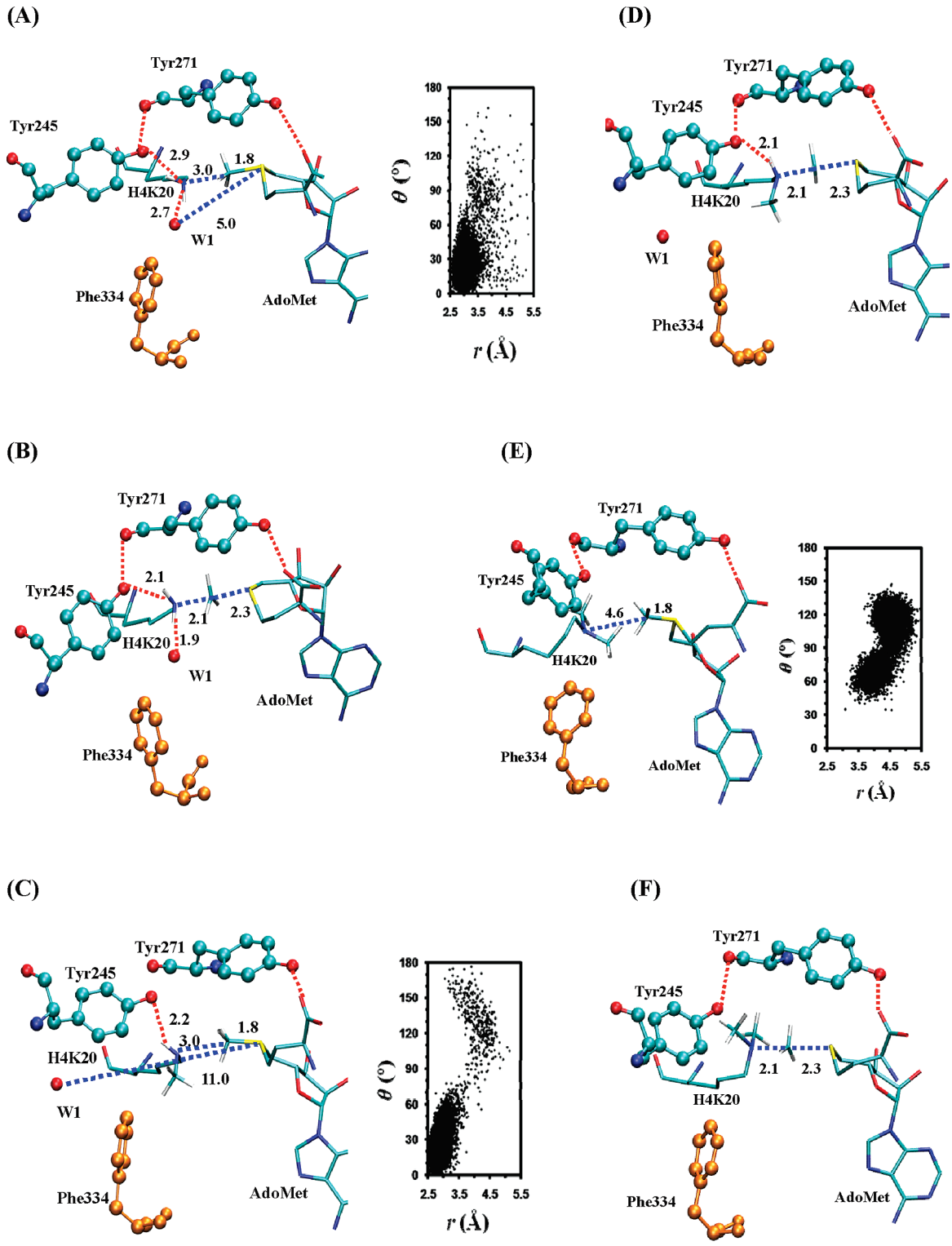

Figure 3. MD results for Y334F. (A) Left: the average active-site structure of the reactant complex for the first methyl transfer. Right: the two-dimensional plot of $r(C_M \cdots N_\zeta)$ and θ distributions of the reactant complex for the first methyl transfer. (B) The average structure near the transition state for the first methyl transfer. (C) Left: the average structure of the reactant complex for the second methyl transfer. Right: the two-dimensional plot of $r(C_M \cdots N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer. (D) The average structure near the transition state for the second methyl transfer. (E) Left: the average structure of the reactant complex for the third methyl transfer. Right: the two-dimensional plot of $r(C_M \cdots N_\zeta)$ and θ distributions of the reactant complex for the third methyl transfer. (F) The average structure near the transition state for the third methyl transfer.

transfer is much higher than that of the first methyl transfer (by as much as 6.5 kcal/mol). Thus, the second methyl transfer is much less efficient compared to the first methyl transfer process, and this is consistent with the experimental findings that SET8 is a monomethylase.[5,7] The earlier single-point MP2/6-31G+G(d,p)/MM calculations led to an average

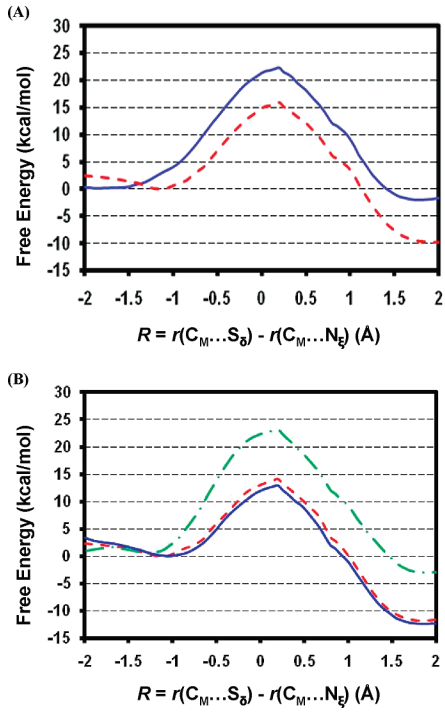

Figure 4. (A) Free energy (potential of mean force) changes for the first and second transfers from AdoMet to H4−K20 and H4−K20me1, respectively, as a function of the reaction coordinate [$R = r(C_M\cdots S_\delta) - r(C_M\cdots N_\zeta)$] in the wild-type SET8. The first methyl transfer: red and dashed line with a free energy barrier of 15.8 kcal/mol. The second methyl transfer: blue and solid line with a free energy barrier of 22.3 kcal/mol (or about 6.5 kcal/mol higher than that of the first methyl transfer). Differences in the free energy barriers may be represented by two energy triplets, (0, Δ_{2-1W}, Δ_{3-1W}) and (Δ_{M-W}, Δ_{2-1M}, Δ_{3-1M}), for the wild-type and mutated enzymes, respectively. For the wild-type enzyme, the second (Δ_{2-1W}) and third (Δ_{3-1W}) parameters are the differences in the free energy barriers between the second and first and between the third and first methyl transfers, respectively. For the mutated enzyme, the first parameter (Δ_{M-W}) is the difference in the free energy barriers for the first methyl transfer in the wild-type and mutant. The second (Δ_{2-1M}) and third (Δ_{3-1M}) parameters are the differences in the free energy barriers between the second and first and between the third and first methyl transfers, respectively, in the mutant. For SET8, (0, Δ_{2-1W}, Δ_{3-1W}) = (0, 6.5, *x*) (*x* indicates the undetermined relative barrier in the energy triplet). (B) The free energy changes for the first, second, and third methyl transfers as a function of the reaction coordinate in the Y334F mutant. The first methyl transfer: red and dashed line with a free energy barrier of 14.1 kcal/mol. The second methyl transfer: blue and solid line with a free energy barrier of 13 kcal/mol. The third methyl transfer: green and dot-dashed line with a free energy barrier of 23.1 kcal/mol. (Δ_{M-W}, Δ_{2-1M}, Δ_{3-1M}) = (−1.7, −1.1, 9).

barrier that is as much as 20 kcal/mol higher for the second methyl transfer than that for the first methyl transfer.[18] The ab initio QM [HF(6-31G*/3-21G*)]/MM free energy simulations for the first and second methyl transfers in two different PKMTs, SET7/9 and Rubisco LSMT, have also been performed previously.[15a] For SET7/9, the free energy barriers for the first and second methyl transfers were calculated to be 22.5 ± 0.5 kcal/mol and 26.2 ± 0.5 kcal/mol, respectively.[15a] Comparing with our earlier results on SET7/9[8a] and taking into account the suggestion that the HF(6-31G*/3-21G*) method may overestimate the barriers by about 3 kcal/mol,[15a] these data indicate that the corrected SCC-DFTB method might underestimate the barriers, although other factors may affect the results as well. However, as mentioned earlier, the relative free energy barriers, instead of the absolute barriers, are expected to be more important in the determination of the product specificity. For SET7/9, the difference for the two methods in the description of the differential free energy barriers is only about 1 kcal/mol.

Figure 4B plots the free energy profiles for the methyl transfers in the Y334F mutant. Unlike the wild-type enzyme, the free energy profiles for both the first and second methyl transfers are rather similar with relatively low barriers. Thus, if the first methyl transfer from AdoMet to the target lysine can be catalyzed by Y334F, the second methyl transfer to monomethyl lysine would also be possible. By contrast, the free energy barrier for the third methyl transfer in Y334F is considerably higher. The results are consistent with the experimental observations that Y334F is a dimethylase[5,7] and support the suggestion[8] that the relative free energy barriers for the methyl transfers are likely to be important energetic factors controlling the product specificity.

Figure 5A and B plot the free energy profiles as a function of the distance between the oxygen atom of W1 and the sulfur atom of AdoMet in the reactant complexes for the first and second methyl transfers, respectively, in the wild-type enzyme. As is evident from Figure 5A and B, the most stable location for W1 is at the active site in these complexes, with a distance of about 5 Å to the sulfur atom, consistent with the average structures observed from the MD simulations (Figure 2A and D). The free energy profiles further show that W1 seems to be held tightly by the active site interactions, as the energetic cost for W1 to move away from the active site in each case is quite high. Figure 5D shows that, for the reactant complex of the second methyl transfer in Y334F, the most stable position for W1 has moved away from the active site (to a location with a distance of about 10−11 Å from the sulfur atom). With the removal of W1, the active site of Y334F becomes less crowded and is able to accommodate the second methyl group on the target lysine. The second methyl transfer can therefore proceed, and the mutant becomes dimethylase (see below).

Discussions

The key question on the product specificity of PKMTs concerns the factor that controls *the methylation state of the product*. This is in contrast with many other investigations on enzyme-catalyzed reactions which concentrate on the effects of enzymes in the reduction of the activation barriers in going from solution to the enzyme active sites. Thus, a convenient reference reaction for understanding the product specificity would be the process involving the first methyl transfer in the wild-type enzyme (see below). The existence of a relatively high barrier for one of the methyl transfer processes may lead to the termination of further methyl addition and therefore determine the product specificity of the enzyme. If none of the three free-energy barriers for the

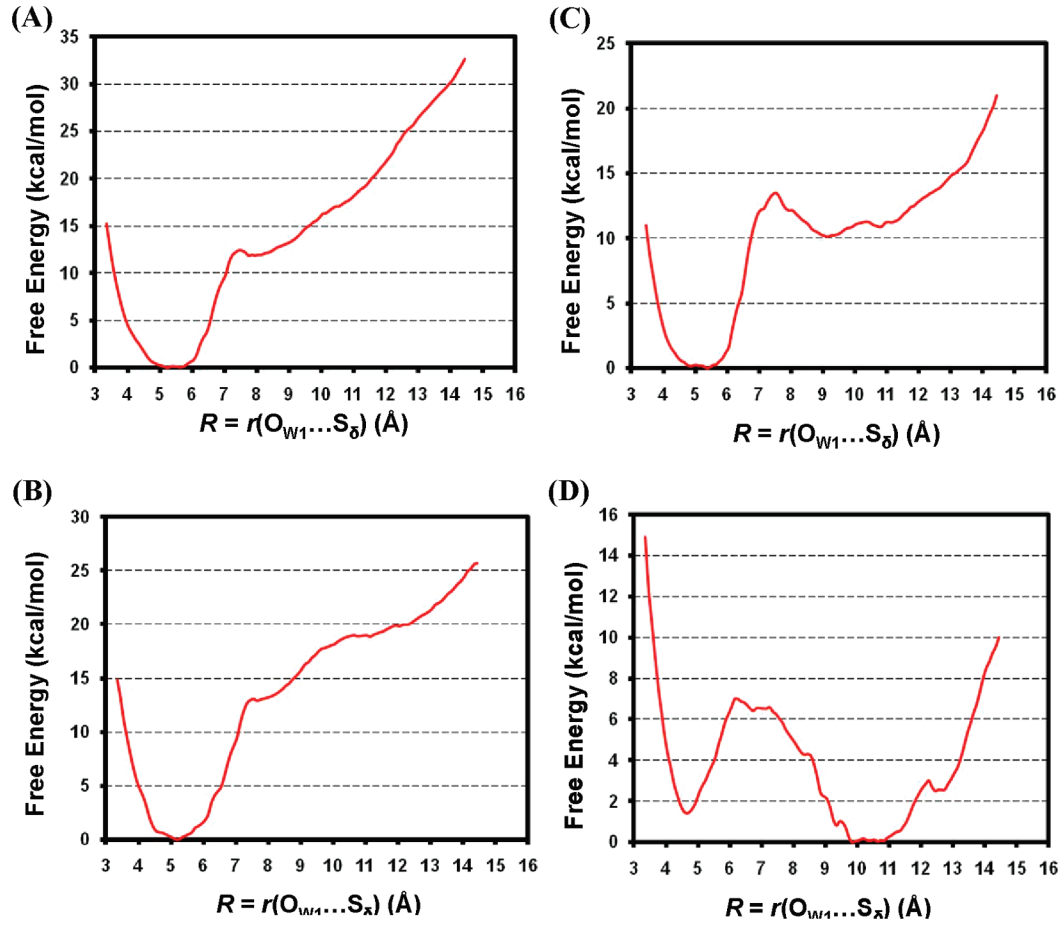

Figure 5. (A) Free energy (potential of mean force) change as a function of the distance between the oxygen atom of the active-site water molecule (W1) and sulfur atom of AdoMet [i.e., $r(O_{w1}\cdots S_\delta)$] in the reactant complex for the first methyl transfer in the wild-type enzyme. W1 is stable in the active site, as the free energy minimum is around 5 Å (i.e., at a similar position to that observed in the MD simulations in Figure 2A). High energy at the longer $r(O_{w1}\cdots S_\delta)$ distance suggests that it would be difficult for W1 to move away from the active site. (B) Free energy change in the reactant complex of the second methyl transfer in the wild-type enzyme. The simulation data show that W1 is also stable in the active site. (C) Free energy change in the reactant complex of the first methyl transfer in Y334F. (D) Free energy change in the reactant complex of the second methyl transfer in Y334F. The most stable position for W1 has changed (now 10−11 Å from the sulfur atom), suggesting that W1 is not in the active site.

methyl transfers is significantly high, the enzyme might be able to catalyze all three methyl transfers and could be a trimethylase (e.g., as in the case of DIM-5; see ref 8b). This proposal is consistent with some previous computational results that suggested that the product specificity is probably to be mainly controlled by the methyl transfer reaction step.[8,15] An alternative explanation of the product specificity is based on the formation of a water channel observed during the MD simulations of PKMTs.[18,19] However, the dramatic increase of the energy barrier from the first to the second methyl transfer in SET8 obtained by the same authors[18] (see above) raises the question concerning the true event that prevents the further methylation in PKMTs (e.g., dimethylation by SET8).

It was proposed in our earlier communication that two different free energy triplets, (0, Δ_{2-1W}, Δ_{3-1W}) and (Δ_{M-W}, Δ_{2-1M}, Δ_{3-1M}) for wild-type and mutated enzymes, respectively, may be used in the description of the product specificity of PKMTs and their mutants.[8b] Here, the free energy barrier for the first methyl transfer in the wild-type enzyme (i.e., the reference reaction) is taken as the zero (for a detailed explanation of the parameters, see Figure 4). For SET8 and its Y334F mutant studied in this work, the corresponding energy triplets can be written as (0, 6.5, x) and (−1.7, −1.1, 9), respectively, which reflect the fact that SET8 is a monomethylase and the mutant a dimethylase (see below). The kinetic study on Y334F[7] suggested that the activation barrier to produce the H4K20me2 product from the H4K20me1 substrate is about 2 kcal/mol higher than the barrier to produce the monomethyl product from the unmodified H4K20 substrate. Figure 4B shows that the second methyl transfer has a free energy barrier that is slightly lower than that for the first methyl transfer (by ∼1 kcal/mol). The simulation data are therefore consistent with the suggestion[7] that methyl lysine reorientation and deprotonation between turnovers may constitute a rate-limiting step in catalysis. However, it should be pointed out that, as far as the product specificity is concerned, the key question is what causes the stop of further methyl addition during histone lysine methylation (which is different from the question concerning the rate-limiting step of the enzyme-catalyzed process). The results of the simulations from the present work on SET8

and our previous studies on DIM-5 and SET7/9[8b] (and certain mutants) suggest that one of the key energetic factors for the specific product specificities of PKMTs is presumably due to a significant increase of the free energy barrier for one of the methyl transfers in the enzymes. Thus, the reason that SET8 is monomethylase is probably due to the fact that the energy barrier for the second methyl transfer is too high and stops further methylation. The Y334→F mutation on SET8 effectively reduces the barrier for the second methyl transfer so that the second methyl transfer can proceed. Since the free energy barrier for the third methyl transfer is very high, the addition of the third methyl group cannot proceed, leading to a dimethylase. Similar arguments may be made for SET7/9 and DIM-5.

It is interesting to note that the free energy data on the ability of SET8 (Y334F) to catalyze the first (first and second) methyl transfer and the inability of SET8 (Y334F) to catalyze the second (third) methyl transfer are already reflected from the MD simulations in Figure 2B and E (Figure 3A, C, E). Similar observations have also been made previously.[8,15,18] Thus, the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful indicators in the prediction of product specificity for PKMTs, although further tests are still necessary to establish the correlations. This result is of importance, because performing the MD simulations is much easier than undertaking the QM/MM free energy simulations. Examination of the structures at the transition states (TSs) in Figures 2 and 3 shows that all these structures are rather similar; e.g., $r(C_M \cdots N_\zeta)$ and $r(S_\delta \cdots C_M)$ are 2.1 and 2.3 Å, respectively, in each of the structures. It is of interest to note these structures are rather close to the TS structures generated earlier from the ab initio QM [HF(6-31G*/3-21G*)]/MM free energy simulations for two different PKMTs, SET7/9 and LSMT.[15a] As discussed earlier, the structures for the corresponding reactant complexes, on the other hand, can be significantly different. Indeed, the structures for the reactant complexes of the first methyl transfer in the wild type (Figure 2A) and the first and second methyl transfers in Y334F (Figure 3A and C, respectively) are rather similar to the corresponding TS structures (Figures 2C and 3B and D, respectively) with a $r(C_M \cdots S_\delta)$ distance of about 4.8 Å. For these cases, a part of the TS stabilization is probably already reflected in the reactant state through the generation of such a TS-like conformation. By contrast, the structures of the reactant complexes for the second methyl transfer in the wild type (Figure 2D) and the third methyl transfer in Y334F (Figure 3E) are significantly distorted from the corresponding TS structure (Figures 2F and 3F, respectively), and the $r(C_M \cdots S_\delta)$ distances are around 6.3 Å. Therefore, additional energetic cost would be required to generate the TS-like structures from these structures, and this could lead to relatively high activation barriers for the corresponding methyl transfers. It should be pointed out that the free energy costs for generating the TS-like structures alone seem not to be sufficient to explain the increases of the barriers for the methyl transfers. Indeed, Figure 2E shows that the free energy cost for producing a structure similar to the one in Figure 2A is approximately 4−5 kcal/mol, while the free energy barrier increases by 6.5 kcal/mol. Therefore, other factors may be involved as well. It is of interest to note from Figure 2A and C that the hydrogen bond distances involving the ε-amino group (with Tyr245 and W1) decrease significantly as the system reaches the transition state. This indicates that the corresponding interactions are strengthening and may play an important role in the transition state stabilization for the methyl transfer as well.

It was proposed that the Phe/Tyr switch may regulate product specificity through altering the affinity of W1, and the dissociation of this water molecule is essential for the multiple methylation process to proceed.[7] Although the structural information on the presence and absence of W1 under different conditions is of considerable interest, question remains concerning how the affinity of W1 would change as a result of the mutation (or methyl addition) and what would be its stability to be located in the active site. The energetic information concerning the stability of W1 at the active site under different conditions is of fundamental importance for the determination of the role of this water molecule in preventing further methlation. Figure 5B shows that W1 is rather stable in the reactant complex of the second methyl transfer in the wild-type enzyme and the free energy cost for its removal is quite high. The high stability is presumably achieved through the interaction involving Tyr334 (Figure 2D). The high energetic cost can make the addition of the second methyl group much more difficult, presumably because the active site becomes too crowded without the removal of W1. This could contribute to the significant increase of the free energy barrier from the first to the second methyl transfer and stop the second methyl addition, although additional simulations are still necessary. Figure 5D shows that, for the reactant complex of the second methyl transfer in Y334F, the most stable location of W1 has changed and W1 moved away from the active site (see also Figure 3D). Thus, this water molecule could not interfere with the methyl transfer process anymore. The mutant is now able to catalyze the second methyl transfer and becomes a dimethlase.

Conclusions

The QM/MM free energy simulations have been performed for the first and second methyl transfers from AdoMet to the target lysine/methyl lysine in SET8 and for the first, second, and third methyl transfers in its Y334F mutant (involving the replacement of the tyrosine/phenylalanine switch residue). The two free energy barriers for the methyl transfers in SET8 and the three free barriers in the mutant obtained from the simulations have been found to be well correlated with the experimental observations on their product specificities. The results indicated that the significant increase of the free energy barrier for the second methyl transfer in SET8 (for the third methyl transfer in Y334F) might stop further methyl addition, and this could be the reason that SET8 (Y334F) is a monomethylase (dimethylase). The results support an earlier suggestion[8b] that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. The QM/MM molecular

dynamics (MD) simulations are also performed on the reactant complexes of the first and second methyl transfers in SET8 and the first, second, and third methyl transfers in Y334F. The results showed that the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers might be used as useful indicators in the prediction of product specificity for PKMTs. The stability of the water molecule at the active site has also been examined on the basis of the free energy simulations. The free energy profiles suggested that the stability of the water molecule at the active site decreases significantly as a result of the Y334→F mutation as well as the methyl addition to the lysine residue. The decrease of the stability of W1 to be located at the active site as a result of the Y334→F mutation is likely to make it easier for the water molecule to dissociate from the active site and create space for further methyl addition.

Acknowledgment. We thank Dr. Hao-Bo Guo, Mr. Mauricio Valenzuela, and Professors Xiaodong Cheng and Jeremy C. Smith for useful discussions and Martin Karplus for a gift of the CHARMM program. This work was supported by the National Science Foundation (Grant number: 0817940 to H.G.). We are also grateful for the computer time from Indiana University TeraGrid Resource.

Supporting Information Available: The estimates of the free energies for formation of the reactive conformations for Y334F and the statistical errors in the potential of mean force simulations. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) (a) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025–1040. (b) Lall, S. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1110–1115. (c) Turner, B. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 110–2.

(2) (a) Jenuwein, T. *FEBS J.* **2006**, *273*, 3121–35. (b) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–49.

(3) (a) Xiao, B.; Wilson, J. R.; Gamblin, S. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 699–705. (b) Cheng, X.; Collins, R. E.; Zhang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267–94.

(4) (a) Zhang, X.; Tamaru, H.; Khan, S. I.; Horton, J. R.; Keefe, L. J.; Selker, E. U.; Cheng, X. *Cell* **2002**, *111*, 117–27. (b) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12*, 177–85.

(5) (a) Couture, J. F.; Collazo, E.; Brunzelle, J. S.; Trievel, R. C. *Genes Dev.* **2005**, *19* (12), 1455–1465. (b) 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.

(6) 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.

(7) Couture, J. F.; Dirk, L. M. A.; Brunzelle, J. S.; Houtz, R. L.; Trievel, R. C. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20659–20664.

(8) (a) Guo, H. B.; Guo, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8797–802. (b) Xu, Q.; Chu, Y.-Z.; Guo, H.-B.; Smith, J. C.; Guo, H. *Chem.—Eur. J.* **2009**, *15*, 12596–12599.

(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(10) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.

(11) (a) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935. (b) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(12) Brooks, C. L.; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843–865.

(13) (a) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268. (b) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.

(14) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(15) (a) Hu, P.; Wang, S.; Zhang, Y. *J. Am. Chem. Soc.* **2008**, *130*, 3806–3813. (b) Hu, P.; Zhang, Y. K. *J. Am. Chem. Soc.* **2006**, *128*, 1272–1278.

(16) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

(17) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(18) Zhang, X. D.; Bruice, T. C. *Biochem.* **2008**, *47*, 6671–6677.

(19) (a) Zhang, X.; Bruice, T. C. *Biochemistry* **2007**, *46*, 5505–14. (b) Zhang, X.; Bruice, T. C. *Biochemistry* **2007**, *46*, 14838–14844. (c) Zhang, X.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5728–5732. (d) Zhang, X. D.; Bruice, T. C. *Biochemistry* **2007**, *46*, 9743–9751. (e) Zhang, X. D.; Bruice, T. C. *Biochemistry* **2008**, *47*, 2743–2748.

CT9006458