CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.200902297

Energy Triplets for Writing Epigenetic Marks: Insights from QM/MM Free-Energy Simulations of Protein Lysine Methyltransferases

Qin Xu,[a] Yu-zhuo Chu,[a] Hao-Bo Guo,[a] Jeremy C. Smith,[a, b] and Hong Guo*[a, b]

The nucleosome is the fundamental building block of eukaryotic chromatin, within which histone proteins play an important role in packaging of DNA.[1] The tails of histone proteins are subject to different post-translational covalent modifications, and these modifications correspond to an important epigenetic mechanism to lead to distinct downstream events in the regulation of chromatin structure and gene expression.[2] One important modification is histone lysine methylation catalyzed by protein lysine methyltransferases (PKMTs).[3–6] The biological consequences of histone lysine methylation (e.g., gene activation and repression) depend on the methylation states of the lysine residue (mono-, di- or tri-methylated; see Figure 1).[7,8] Therefore, it is of fundamental importance to understand why different PKMTs have their unique ability to direct specific degrees of lysine methylation which is called product specificity. Such knowledge may have important implications for developing strategies in the manipulation of the signaling properties.

In this Communication, the free-energy profiles are obtained from quantum mechanical/molecular mechanical (QM/MM) free-energy simulations for the first, second and third methyl transfers in DIM-5 (a trimethylase) as well as in some of its mutants with different product specificity. The free-energy profile for the third methyl transfer in SET7/9 (a mono-methylase) is also obtained and compared with the data published earlier.[9] It is found that in each case the three free-energy barriers are well correlated with experimentally observed product specificity. The results of the simulations suggest that the relative efficiencies of the chemical steps involving the three methyl transfers in PKMTs from S-adenosyl-L-methionine (AdoMet) to the ε-amino group of the target lysine may determine, at least in some cases, how the epigenetic marks of lysine methylation are written. Two different energy triplets are proposed as important parameters for the prediction of product specificity.

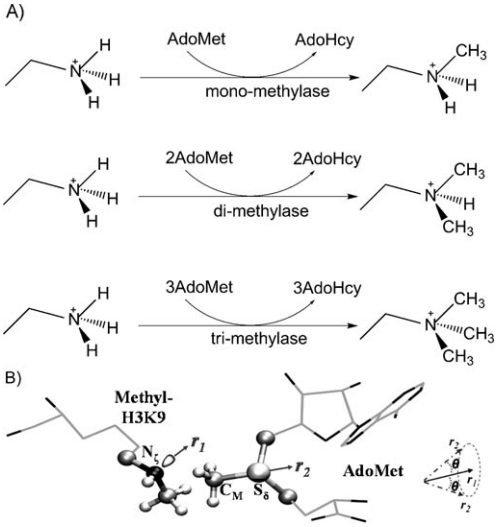

Figure 1. A) Mono-, di- or tri-methylation of Lys. B) The reaction coordinate: $R = r(C_M \cdots S_\delta) - r(C_M \cdots N_\varepsilon)$. The parameters for monitoring the orientation of AdoMet and lysine/methyl lysine are $r(C_M \cdots N_\varepsilon)$ and θ. θ is defined as the angle between the two vectors r_1 (the direction of the lone pair of electrons) and r_2 (the direction of $C_M - S_\delta$ bond pointing from C_M to S_δ).

The free-energy profiles for the first, second and third methyl transfers are plotted in Figure 2A for DIM-5 as a function of the reaction coordinate. It is of interest to note that the free-energy barriers are rather similar. Thus, if the first methyl transfer from AdoMet to the target lysine can

[a] Dr. Q. Xu, Y.-z. Chu, Dr. H.-B. Guo, Prof. J. C. Smith, Prof. H. Guo
Department of Biochemistry and Cellular and Molecular Biology
University of Tennessee
Knoxville, TN 37996 (USA)
Fax: (+1) 865-974-6306
E-mail: hguo1@utk.edu

[b] Prof. J. C. Smith, Prof. H. Guo
UT/ORNL Center for Molecular Biophysics
Oak Ridge National Laboratory
Oak Ridge TN 37831 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200902297 and contains a comparison of the results of the B3LYP/6-31G** and corrected SCC-DFTB calculations for the methyl transfer in a model system, a detailed description of the methods, and an estimate of the free energy for formation of the reactive conformations..

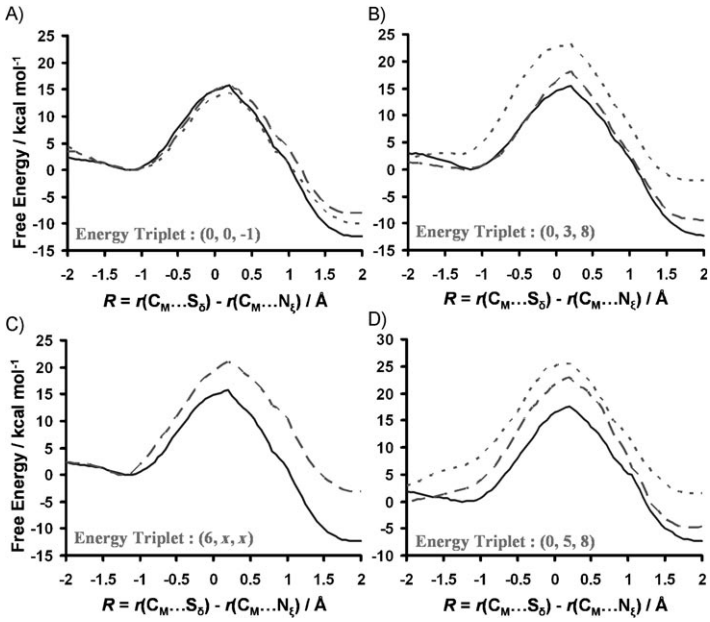

Figure 2. Free-energy profiles: First methyl transfer (——, or otherwise noted); second methyl transfer -----); third methyl transfer (•••••). Energy triplets (0, Δ_{2-1W}, Δ_{3-1W}) and (Δ_{M-W}, Δ_{2-1M}, Δ_{3-1M}) are defined for wild-type and mutated enzyme. For wild-type enzyme the second (Δ_{2-1W}) 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. A) Wild-type DIM-5. B) F281Y. C) Comparison of the free-energy profiles for the first methyl transfer in wild-type and F281W. Wild-type (——); F281W (-----, *x* indicates the undetermined relative barriers in the energy triplet). D) Comparison of the free-energy profile of the third methyl transfer with the first and second methyl transfers in SET7/9.[9]

be catalyzed by DIM-5, the second and third methyl transfers would also be possible. The results are therefore consistent with the fact that DIM-5 is a tri-methylase.[10,11] For F281Y (Figure 2B), the free-energy barrier increases from the first to second methyl transfer (by $\approx$3 kcal mol^{-1}) and from the first to the third methyl transfer (by $\approx$8 kcal mol^{-1}). Thus, unlike the wild-type enzyme, the ability of F281Y to catalyze methyl transfer decreases as more methyl groups are added to the target lysine. Presumably, much of the product of mono- and/or di-methylation would already be released before the third methyl group could be added, consistent with the experimental findings that F281Y is a mono-/di-methylase.[11] The results suggest that the relative free-energy barriers for the methyl transfers may be important energetic factors controlling the product specificity. Thus, we design two 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 enzyme, respectively, for understanding and explaining the product specificity (see Figure 2 for details).

To examine whether the free-energy approach and energy triplets can go beyond the simple Phe→Tyr mutation in DIM-5, the free-energy profile was obtained for the first methyl transfer in F281W which did not show any DIM-5 MTase activity.[11] Figure 2C compares the free-energy profiles for the first methyl transfer in wild-type and F281W. Consistent with the experimental observation,[11] the free-energy barrier increases significantly as a result of F281→W mutation. Figure 2D compares the free-energy profile for the third methyl transfer in SET7/9 with those of the first and second methyl transfers obtained earlier.[9] In agreement with the fact that SET7/9 is a mono-methylase, this enzyme has an energy triplet of (0, 5, 8) with much higher barriers for the second and third methyl transfers. The present proposal on the importance of energy triplets in understanding product specificity is consistent with previous suggestions that the methyl transfers are the rate-limiting steps in the lysine methylation[12] and that local arrangements of the active sites are important.[13] Interestingly, the methyl transfer was also proposed to be the rate limiting in the reaction catalyzed by catechol *O*-methyltransferase.[14]

To understand the reason for the existence of different barriers for the methyl transfers, the average active-site structures of the reactant complexes for the third methyl transfers in DIM-5 and F281Y are given in Figure 3A,C, respectively. The structures near the transition states (TSs) for the corresponding methyl transfers are also plotted (Figure 3B,D). As is evident from Figure 3A, the lone pair of electrons on N_ζ of the di-methyl lysine is well aligned with the methyl group of AdoMet, and this is further demonstrated by the large population of the structures with relatively short $r(C_M \cdots N_\zeta)$ distances and small θ angles. Figure 3A,B show that the structure of the reactant complex for wild-type is rather similar to the structure near TS. This is in contrast to the case of F281Y where the reactant structure (Figure 3C) is significantly distorted from the corresponding TS structure (Figure 3D). Thus, one of the reasons for the existence of the relatively low free-energy barrier is likely owed to the fact that a part of TS stabilization is already reflected on the reactant state through the generation of the TS-like conformation. This is consistent with the previous analysis based on chorismate mutase[15] that the catalytic effect in the formation of the reactive reactant conformation is simply the result of TS stabilization rather than a ground state effect. An estimate of this effect in the case of DIM-5 and F281Y is given in the Supporting Information. Notably, in Figure 3 the active-site water (W1) was stabilized at the active site of F281Y without dissociation, whereas it was pushed away in the wild-type enzyme. This observation is consistent with the experimental results on SET8[13] and suggests the importance of releasing the active-site water molecule for multiple methyl addition in DIM-5.

The key question concerning the product specificity is to understand the factors that control the methylation state of the product (i.e., the number of methyl groups added). The existence of a good correlation between the relative free-energy barriers and experimentally observed product specificity/activity strongly suggests that the relative efficiencies of the methyl transfers may determine, at least in some cases,

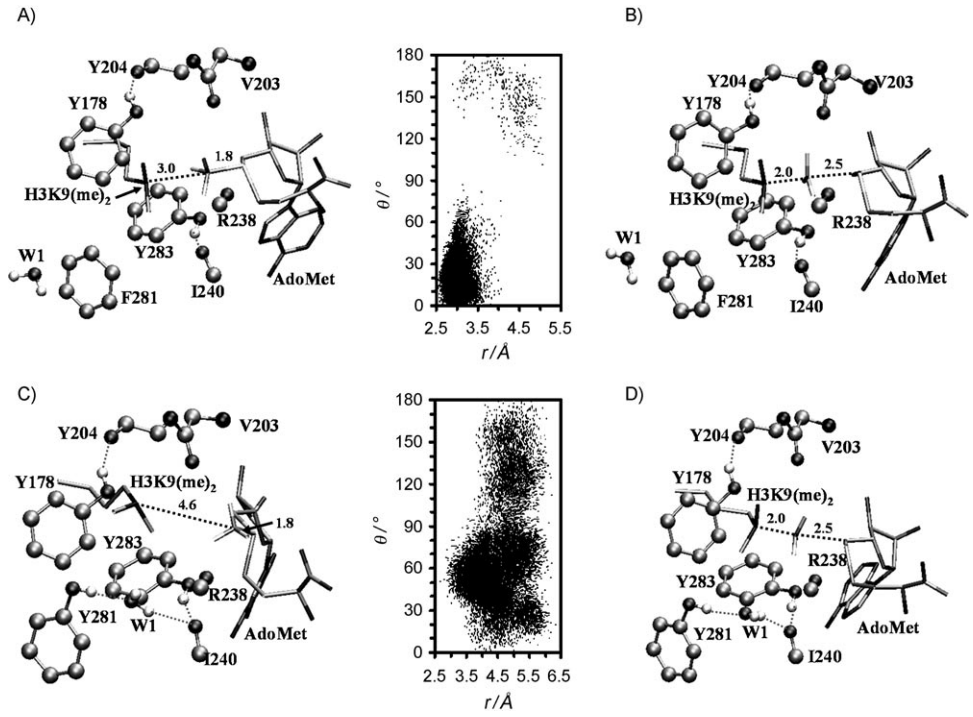

Figure 3. A) Average structure of the reactant complex for the third methyl transfer in DIM-5 along with $r(C_M\cdots N_\zeta)$ and θ distributions. The enzyme is shown in balls and sticks, and AdoMet and H3K9(me)$_2$ in sticks. The hydrogen bonds and distances to the methyl group are indicated by dotted lines. The hydrogen atoms of the methyl groups on N_ζ are not shown for clarity. Oxygen and nitrogen atoms: dark grey; Carbon atom: light grey; Hydrogen atom: white. B) Average structure near TS for the third methyl transfer in DIM-5. C) Average structure of the reactant complex for the third methyl transfer in F281Y along with $r(C_M\cdots N_\zeta)$ and θ distributions. D) Average structure near TS for the third methyl transfer in F281Y.

the product specificity of PKMTs and that the energy triplets proposed in this work may be used as important parameters in the prediction of the product specificity.

Methods

The QM/MM MD and free-energy (potential of mean force) simulations were applied to characterize the methyl transfers by using the CHARMM program.[16] AdoMet/AdoHcy and lysine/methyl-lysine side chains were treated by QM and the rest of the system by MM. The link-atom approach[17] was applied to separate the QM and MM regions. The SCC-DFTB[18] method was used for the QM region with an empirical correction (see the Supporting Information). The all-hydrogen potential function (PARAM27)[19] was used for the MM atoms. A modified TIP3P water model[20] was employed for the solvent. The stochastic boundary molecular dynamics method[21] was used for the QM/MM MD and free-energy simulations. The initial coordinates for the reactant complexes were based on the crystallographic complex (PDB code: 1PEG) of DIM-5.[10,11] After 1.5 ns QM/MM MD simulations were carried out for each of the reactant complexes of the methyl transfers, the umbrella sampling method[22] along with the Weighted Histogram Analysis Method (WHAM)[23] was applied. The reaction coordinate was defined as a linear combination of $r(C_M-N_\zeta)$ and $r(C_M-S_\delta)$ [i.e., $R=r(C_M-S_\delta)-r(C_M-N_\zeta)$]. For a complete description of the methods used and their suitability, see the Supporting Information.

Acknowledgements

We thank Prof. Xiaodong Cheng for useful discussions. This work was supported by the National Science Foundation to H.G. (Grant number: 0817940) and in part by a DOE Laboratory-Directed Research and Development award to J.C.S.

Keywords: density functional calculations · enzyme catalysis · free-energy simulations · histone lysine methylation · quantum chemistry

[1] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251–260.
[2] S. Lall, *Nat. Struct. Mol. Biol.* **2007**, *14*, 1110–1115.
[3] T. Jenuwein, *FEBS J.* **2006**, *273*, 3121–3135.
[4] C. Martin, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–849.
[5] A. J. Ruthenburg, C. D. Allis, J. Wysocka, *Mol. Cell* **2007**, *25*, 15–30.
[6] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025–1040.
[7] X. Cheng, R. E. Collins, X. Zhang, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267–294.
[8] B. Xiao, J. R. Wilson, S. J. Gamblin, *Curr. Opin. Struct. Biol.* **2003**, *13*, 699–705.
[9] H. B. Guo, H. Guo, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8797–8802.
[10] X. Zhang, H. Tamaru, S. I. Khan, J. R. Horton, L. J. Keefe, E. U. Selker, X. Cheng, *Cell* **2002**, *111*, 117–127.
[11] X. Zhang, Z. Yang, S. I. Khan, J. R. Horton, H. Tamaru, E. U. Selker, X. Cheng, *Mol. Cell* **2003**, *12*, 177–185.
[12] P. Hu, S. Wang, Y. Zhang, *J. Am. Chem. Soc.* **2008**, *130*, 3806–3813.
[13] J. F. Couture, L. M. A. Dirk, J. S. Brunzelle, R. L. Houtz, R. C. Trievel, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20659–20664.
[14] M. F. Hegazi, R. T. Borchardt, R. L. Schowen, *J. Am. Chem. Soc.* **1979**, *101*, 4359–4365.
[15] M. Strajbl, A. Shurki, M. Kato, A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 10228–10237.
[16] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
[17] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700–733.
[18] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, *J. Phys. Chem. B* **2001**, *105*, 569–585.
[19] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-

McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[21] C. L. Brooks, A. Brunger, M. Karplus, *Biopolymers* **1985**, *24*, 843–865.

[22] G. M. Torrie, J. P. Valleau, *Chem. Phys. Lett.* **1974**, *28*, 578–581.

[23] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J. Comput. Chem.* **1992**, *13*, 1011–1021.

Received: August 19, 2009
Published online: October 30, 2009

