Cascading proton transfers are a hallmark of the catalytic mechanism of SAM-dependent methyltransferases

Li Na Zhao\textsuperscript{1} and Philipp Kaldis\textsuperscript{1,2}

\textsuperscript{1} Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
\textsuperscript{2} Department of Clinical Sciences, Lund University, Clinical Research Center (CRC), Malmö, Sweden

Correspondence
L. N. Zhao, Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
Tel: +65 9866 9611
E-mail: lina.oahz@gmail.com

(Received 1 January 2020, revised 20 April 2020, accepted 22 April 2020, available online 16 May 2020)
doi:10.1002/1873-3468.13799
Edited by Peter Brzezinski

The SET (suppressor of variegation 3–9, enhancer of zeste and trithorax) domain is a ~130 amino acid, evolutionarily well-conserved motif, which can be found in proteins/enzymes from yeast to mammals, and even some bacteria and viruses [1]. In 1998, there were 368 SET domain proteins [2] annotated, and nowadays, we count up to 53 403 proteins in the SMART’s nrdb database across all kingdoms of life. These SET domain proteins are key to epigenetic regulation of gene activation and silencing in eukaryotic organisms and affect cellular proliferation and differentiation [3,4].

The SET domain-containing enzymes use the cofactor S-adenosyl-l-methionine (SAM; AdoMet) as a methyl group donor for the methylation of protein substrates. SET domain lysine methyltransferases catalyze methylation of lysine residues in histone H3 and H4, linker histone H1B, and nonhistone substrates in a context-dependent manner [5], which are essential for regulation of post-translational modifications (PTM), replication, DNA repair, DNA recombination, cell cycle control, and metabolism [6,7]. The most common residues of histone lysine methylation are H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20 [8]. The methylation of H3K4, H3K36, and H3K79 is usually associated with transcriptional active genes, while H3K9, H3K27, H4K20, and H1bK26 frequently correlate with a silent state of gene expression [9,10].

The PRDM (PRDI-BF1 and RIZ homology domain containing) protein family, a distinct class of SET domain proteins, plays an important role in cell differentiation. They are dysregulated in human diseases, most notably in hematological malignancies and cancers [4,11,12]. The PRDM family, sometimes described for a lack of intrinsic lysine methyltransferase activity (may be due to the difficulty to predict their

Abbreviations
EVB, empirical valence bond; LRF, local reaction field; MCPT, Monte Carlo proton transfer; PDLD, protein dipole Langevin dipole; PTM, post-translational modifications; SAM, S-adenosyl-l-methionine; PRDM, PRDI-BF1 and RIZ homology domain containing.
substrates), consists of 17 proteins in human and only a few [4,13,14] display well-defined enzymatic activities, which leaves a knowledge gap about their activity especially at molecular level [15]. For example, PRDM15 transcriptionally regulates the WNT and MAPK-ERK signaling pathway [15], but substrates and enzymatic activity have not been identified.

PRDM9, as one of the few PRDM methyltransferases with a determined 3-dimensional structure, displays high activity in catalyzing mono-, di-, and tri-methylations of the H3K4 mark as well as H3K36 [16]. Therefore, PRDM9 serves as a good representative of the PRDM family to study the SAM-dependent methyltransferase reaction mechanism in molecular detail.

The SAM-dependent lysine methyltransferases employ a conventional $S_n2$ mechanism in which the methyl group of SAM is transferred to the $N_e$ of the methyl lysine. It has been suggested that the reaction requires an in-line attack of the nucleophile toward the C atom of the methyl group (electrophile) accompanied by the sulfur atom as leaving group. During the reaction, with the inversion of the chiral center (Walden Inversion) of the methyl group, a planar arrangement of the carbon atom with the remaining three hydrogens is formed as the transition state [17,18].

Prior to the methyl transfer, the methyl lysine in the substrate must be deprotonated, which is an important step in the catalytic cycle [19]. As it is well known that the $pK_a$ of lysine side chain is around 10.5 [20], the substrate requires a mechanism to deprotonate the lysine for the following methyl transfer [19]. Due to the absence of functional groups at the active site to act as a general base for the deprotonation of the substrate lysine, two hypotheses were proposed earlier: (a) only a few lysine side chains that are deprotonated at the pH of the nucleus are available to be methylated [21]; or (b) the substrate lysine side chain has already been deprotonated when it binds to SET domain-containing HKMT in the presence of AdoMet [22]. Since neither of these hypotheses are attractive, we were looking for alternatives and focused on three candidate tyrosines (Y357, Y276, and Y341) at active site that could potentially fulfill the deprotonation function. Consensus studies indicate (including our study below) that the mutation Y357F abolishes PRDM9 enzyme activity and did not affect the methyl transfer stage [13,19]. Thus, Y357 could be potentially responsible for the deprotonation of the methyl lysine after all, and Y276 and Y341 may contribute further to the deprotonation of mono-methylated lysine and dimethylated lysine.

Experimental observations lead to the assumption that Tyr357, a well-conserved residue in PRDMs [23], provides the hydrogen bond of O-CH from SAM-dependent methyltransferases (O from Tyr357 in PRDM9), which greatly augment SAM binding affinity and transition-state stability [24]. This encourages computational studies into how this residue contributes to the catalytic mechanism. Actually, the catalytic relevance of Tyr357 cannot be established without calculating the free energy of the chemical reaction that would occur in the corresponding proton transfer and methyl transfer steps. However, in the past extensive computational studies have been dedicated especially to the methyl-transfer step [22,25–27], while no study has been done on the proton transfer step. In our study, we integrated the proton transfer and methyl transfer step and illustrated how the deprotonation would take place. In addition, the ‘Y357F mutation abolishes enzyme activities’ [13] provides us an important clue in the investigation of the PRDM9 reaction mechanism. Therefore, we have examined comprehensively how Y357 contributes to the proton transfer reaction.

Since the reaction mechanism of enzymes is difficult to study with experiments due to the short time frame of the reaction (the transition state is unstable) and due to that mutations often abolish enzymatic activity, we chose an array of computational efforts to demonstrate that (a) the catalytic proton is in a proton-transfer-ready (here we referred it as a ‘tunneling ready’) reactant state, and Tyr357 bridges the two proton-transfer-ready wells (shared low-barrier H-bond); (b) approaching of SAM triggers the cascading proton transfer effect; and (c) the proton is likely to be released to the bulk solvent accompanied by the release of the product. Our free energy calculations are consistent with experimental kinetic data, suggesting that Tyr357 bridges tunneling ready states for the proton transfer and this is essential for the deprotonation of the substrate methyl lysine.

**Materials and methods**

**Empirical valence bond calculations**

The initial structure of the PRDM9 was obtained from the protein data bank (PDB ID: 4C1Q) with the crystal waters retained for our simulation. H3K4me2 (Kme2) was tailored to H3K4me0 (Kme0) and H3K4me1 (Kme1) for our different simulation systems, and the product S-adenosyl homocysteine (SAH; AdoHcy) was converted back to the substrate SAM using OpenEye Toolkits [28] for the study of the enzyme reaction. The empirical valence bond approach [29–31] was used to simulate the proton and methyl transfer reaction mechanism as depicted in Fig. 1.
The general procedure has been detailed in Supporting Information and also in our previous work [32–34]. The general polarizable ENZYMIX force field [35] within MOLARIS-XG software was used to represent all atoms. The water molecules are added into the system within the radius of 12 Å from reaction center, and then, a large radius (20 Å) for region II is specified, followed by a bulk continuum. The residues falling beyond this sphere were treated at their modeled positions. Local reaction field (LRF) was used for the long-range effects [36]. The system was relaxed for 50 ns and equilibrated in a physiological-mimic environment using the surface constrained all-atom protein/solvent model [37] within MOLARIS-XG. Note that 300 K was used as the standard temperature for simulation and the calculations of activation free energies. The side chain of ionizable residues was evaluated and calculated by coarse-grained Monte Carlo proton transfer (MCPT) module within MOLARIS-XG. The final configurations were used for the subsequent EVB studies. Free energy perturbation and umbrella sampling [38] have been used in the EVB free energy calculation by 51 frames of 100 ps each. The protonation state of key ionizable residues is provided in Supporting Information as Table TS2. The partial charges and EVB parameters are listed in Supporting Information as Table TS3–TS8.

Binding free energy calculations

The binding free energies of the substrate with WT, Y276F, Y341F and Y357 were calculated by the semi-macroscopic version of the protein dipole Langevin dipole (PDLD) with the linear response approximation (PDLD-S/LRA), in which the water molecules are represented by the Langevin dipoles. This method has been established to provide a reliable estimation of the binding free energies [39]. Additionally, more careful consideration of the dielectric response has been used for the different states of ε-amino group and the cofactor SAM as well as the ionizable residue. Thus, we estimated the dielectric constant of four for neutral protein, and an effective dielectric constant of 60 for the charge-charge interaction between the substrate, cofactor, and the side chains of the ionizable residues. The justification of our treatment concept has been discussed in great detail in Ref. [40].

pKa calculations

The PDLD model [29] has been used to calculate the pKa value of the active site tyrosine residues. 2 ps MD relaxation was done in the microscopic framework of the linear response approximation for each calculation. The configurations of both neutral and ionized states of the protein were generated. The standard approach has been developed and used for decades in Dr. Warshel’s group [41], and was very effective in calculation of the pKa values of residues in proteins [41,42]. All the calculations were done using POLARIS module in MOLARIS-XG [35].

Water flooding

The water flooding (WF) has been developed to accelerate the insertion of water by highly saturating the protein with an excess number of internal water molecules, and then, the postprocessing Monte Carlo (MC) strategy combined with the linear response approximation and the linear interaction energy (LIE) was applied to only keep the most likely configurations of the internal water molecules. While the development and the validation of the water flooding model have been described extensively in Ref. [43,44], we have given a brief introduction of the water flooding method in Supporting Information, and here, we give the key parameters used in this study. SCAAS surface constraints and the local reaction field (LRF) long-range treatment were used in our simulation. About 10 000 steps of minimization followed by 200 ps MD relaxation were done on our initial structure, using the polarizable ENZYMIX force field with a time step of 1.0 fs. These structures were then used for WF simulations. During the WF simulations, a spherical hard wall was placed so that the inside water molecules could not escape and the outside water molecules could not enter. The radius of spherical hard wall for cavity was 6.0 Å.

Gaussian software aided calculations

The direct calculations of the free energy profile in aqueous solutions were done using GAUSSIAN09 software [45] with M062X method at 6-31+G* level. The polarizable continuum model (PCM) was used for all reaction trajectory calculations. For the active site at ground, intermediate, and product state, the structures were constructed based on the configuration from 4C1Q.pdb. The transition-state structure is obtained from the TS (Berny) optimization, which shows only one imaginary frequency. The IRC path was calculated starting from the optimized TS, and the energy maximum on the path was located as the transition state. The reactant and product were characterized as having no imaginary frequency. Note that our QM region consists of hydroxide, Tyr357 side chain, and ε-amino group of methyl lysine for proton transfer and the sulfonium and ε-amino group of methyl lysine for methyl transfer (see Fig. 1).

Results and Discussion

Proton transfer and methyl transfer mechanism in water

In principal, unprotonated Asp, Glu, and His as well as water can act as a general base for the proton abstraction from the substrate Lys [46]. In water, the
pKa values for the side chain of Asp, Glu, His, and Lys are 3.65, 4.25, 6.0, and 10.53, respectively [20]. The pKa value of water is estimated to be 14. Hence, the free energy for the proton transfer forms a Lys residue to the general bases, which were calculated based on $\Delta G_{PT} = 2.3RT [pKa_{[base]} - pKa_{[Lys]}]$ and given in Table 1, whereas the reaction coordinate for the methyl group transfers from SAM to different Lys methylated states (Kme0, Kme1, and Kme2), which were calculated at M062X/6-31+G* level with Gaussian and listed in Table 1 and Fig. S2. Additionally, the reaction profile in water listed in Table 1 provides us the EVB parameters to calibrate the reactions in the protein [30,31]. The methyl transfer in water is favorable with the activation energies for the first, second, and the third methyl transfer being 13.8 kcal·mol$^{-1}$. From Fig. S2, we can deduce that the free energy of the Kme0 decreases which provides a rough but clear indication of the whole reaction pathway. Since the methyl transfer share the same reaction profile and all the three methyl transfer share the same mechanism (see Fig. 1 and Fig. S2), it appears that the H3K4 mono-methylation represents reasonably well the entire reaction profile of PRDM9.

### The methyl transfer mechanism in PRDM9

The same parameters we have obtained from the methyl transfer reaction in water were calibrated in protein, and we calculated 8.09 ± 0.6 kcal·mol$^{-1}$ as reaction free energy for the methyl transfer step in PRDM9 (see Table 2). The free energy for the mutants of catalytic essential residues Y276F, Y341F, and Y357F was close to the WT, which indicates that these mutants did not affect the $k_{cat}$ value of the methyl transfer. The binding energy calculation of the

---

**Table 1.** Calculated free energy for the catalytic reaction of Kme0 methylation. The $\Delta G_{obs}$ is calculated based on the experimental value $k_{cat} = 19 000$ h$^{-1}$ using transition-state theory [16].

| Kme0 | Candidate residues for possible proton transfer | $\Delta G_{PT}$ | $\Delta G_{PT}^{\uparrow}$ | $\Delta G_{MT}$ | $\Delta G_{bind}$ |
|------|-----------------------------------------------|----------------|------------------------|----------------|----------------|
| Water | Asp                                          | 9.4           | 13.8                   | 23.2          | 16.6          |
|       | Glu                                          | 8.58          | 13.8                   | 22.38         |               |
|       | His                                          | 6.17          | 13.8                   | 19.97         |               |
|       | H$_2$O                                       | 16.85         | 13.8                   | 30.65         |               |
| PRDM9 |                                              |               |                        |               |               |

$\Delta G_{PT}$, calculated reaction free energy for the proton transfer; $\Delta G_{PT}^{\uparrow}$, calculated activation free energy for the methyl transfer; $\Delta G_{MT}$, calculated total activation energy for reaction; $\Delta G_{bind}$, experimental observed value.

**Table 2.** Calculated free energy (kcal·mol$^{-1}$) for the methyl transfer reaction in wild-type (WT) PRDM9 and mutants as well as substrate binding energy.

|                   | $\Delta G_{MT}$ | $\Delta G_{PT}^{\uparrow}$ | $\Delta G_{bind}$ |
|-------------------|----------------|--------------------------|------------------|
| Kme0              |                |                          |                  |
| Water             | 13.8           | –27.61                   |                  |
| WT                | 8.09           | –14.51                   | –7.11            |
| Y276F             | 8.28           | –17.03                   | –6.47            |
| Y341F             | 7.12           | –18.22                   | –7.94            |
| Y357F             | 8.05           | –16.54                   | –7.6             |

$\Delta G_{MT}$, calculated activation free energy for the methyl transfer step; $\Delta G_{PT}^{\uparrow}$, calculated reaction free energy of the methyl transfer; $\Delta G_{bind}$, calculated binding energy.
substrate as well as of the cofactor in wild-type PRDM9 and the mutants indicates that none of these mutants significantly change the binding energy; namely, the effect of these mutants toward the $K_M$ is negligible. Overall, our calculation indicates that the Y357F mutation does not affect the methyl transfer in terms of $k_{cat}$ and $K_M$.

### Proton transfer in PRDM9

**Deprotonated before entering the active site?**

How the proton transfer in PRDM9 happens has not yet been determined. One possibility is that the substrate lysine side chain has already been deprotonated before it binds to the SET domain-containing HKMT in the presence of AdoMet [22]. From our Table 1, we deduce that the energy after the proton transfer ($\Delta G^{PT}$) is 9.4, 8.58, 6.17, and 16.85 if Asp, Glu, His, and H$_2$O abstract the proton, respectively. Adding the 8.09 kcal mol$^{-1}$, the total energy barrier for the reaction results in 17.49 kcal mol$^{-1}$, 16.67 kcal mol$^{-1}$, 14.26 kcal mol$^{-1}$ and 24.94 kcal mol$^{-1}$, respectively. Based on the experimentally observed energy barrier 16.6 kcal mol$^{-1}$ [16], there is a possibility that the substrate lysine side chain has already been deprotonated by a general base like Asp, Glu, and His before it enters the catalytic site. As a consequence, experimental observations of the H3K4 in the deprotonated state should be evident in some of the previous studies. So far, structural or biochemical data do not support this possibility, which makes it unlikely that the lysine is deprotonated before binding to PRDM9.

**$pK_a$ value of Tyr357**

Another possibility is that Tyr357, as a well-conserved residue in PKMTs and PRDMs, functions in the deprotonation of the substrate lysine. The reason, as has been discussed in the SET7/9 case study [19], is due to that Y357F mutation abolishes the enzyme activity [13], but the Y357F mutation does not affect the methyl transfer reaction barrier (which is evident in our study); hence, it could play a critical role in the proton transfer step. However, in the SET7/9 case study [19], the energy barrier for the methyl transfer stage has already accounted for the experimental observed value in order for the reaction to take place, which means the energy cost for the proton transfer should not accrue any more energy barrier.

Whether Tyr357 can deprotonate Lys, or mono-, even dimethylated lysine is unclear and requires further study. Additionally, there are three tyrosines (Tyr276, Tyr341, and Tyr357) at the active site, which potentially contribute to the proton transfer with the possibility that Tyr357 is directly related to the deprotonated lysine, and Tyr276 (or Tyr341) is responsible for the further deprotonation of the mono-methylated lysine, and then, Tyr341 (or Tyr276) takes charges of the deprotonation of the dimethylated lysine. Furthermore, mutation of the Tyr276 or Tyr341 may be corresponding to different product specificity of the methyltransferase.

In order to study the roles of the tyrosines in the deprotonation of the methyl lysine, it is essential to estimate the $pK_a$ value of these residues. It is well known that the $pK_a$ values of Tyr276, Tyr341, and Tyr357 in protein usually are very difficult to establish due to the internal structure of the enzyme, especially when the local environment changes upon substrate binding [47, 48].

Here, we used Monte Carlo proton transfer (MCPT) and the semimicroscopic version protein dipoles Langevin dipoles (PDDL) model to calculate the $pK_a$ value of the three tyrosines with the substrate Lys in three different states (Kme0, Kme1, Kme2). Note that the Kme1 has three possible configurations with respect to the surrounding tyrosines; here, there are denoted as c1, c2, and c3. In terms of the ‘true’ protein dielectric constant ($\tau$), a lot of work, justification, and discussion have been published in Ref. [41]. Studies have revealed that in the active sites or the sites around the ionizable residues, the optimal $\tau$ is around 8. Here, we calculated the $pK_a$ value of the Tyr357 along the effective dielectric constant ($\epsilon_{eff}$), at Kme0 with SAM (w/SAM) and without (w/o) at different methylated states (Kme0, Kme1 in three different configurations, and Kme2). The results are depicted in Fig. 2. The apparent $pK_a$ values at $\epsilon_{eff} = 8$ are given in Supporting Information as Table S1. We found that the acquired $pK_a$ value of tyrosine inside the protein was decreased significantly (see Figs 2 and S4–S7 in Supporting Information). The $pK_a$ of Tyr357 is generally found to be below 10 at the inside of the protein for the entire range of $\epsilon_{eff}$. The very low $pK_a$ value of Tyr in the presence of the ionized Lys waiting to be deprotonated makes the tyrosine residue an ideal proton acceptor. Furthermore, we also calculated the $pK_a$ values of the tyrosines when SAM is not present. We found that even without SAM, the $pK_a$ values of the tyrosines are also decreased, but far less than in the presence of SAM. For a more realistic view, we consider the average $pK_a$ value obtained by the consistent calculations of $pK_a$ at the semimicroscopic version with the effective $\epsilon$ value. From Fig. 2, we conclude that in all the states we examined (Kme0: see Fig. 2;
Kme1 and Kme2: see Figs S4–S7) the presence of SAM decreased the pKa value low enough for tyrosine to function as an effective proton acceptor.

**Kme0⋯Tyr357⋯OH− proton relay**

In the presence of SAM, the decreased pKa value of the tyrosine makes it feasible to abstract the proton from methyl lysine. By analyzing the high-resolution X-ray structure of the PRDM9 [13], we observed that there is no nearby base residues available for a proton relay network. For the nearby Val319 backbone, there is lack of downstream acceptor to relay the proton out. Therefore, the water-mediated proton transfer mechanism becomes the only feasible possibility for the deprotonation of the methyl lysine. The available PRDM9 X-ray crystal structure does not give us information about the presence of water molecules at the active site. Structural analysis indicates that the accessibility to solvent was also limited due to the presence of SAM. Hence, the approaching and binding or even the release of the SAM/SAH becomes the main source triggering the penetration of water into the catalytic site.

The water, in and around the protein, is important to quantify the function of enzymes. The consideration of the water molecules, particularly these at the active site, is challenging using computational approaches due to the insertion of water molecules that may lead to a high activation barrier and demand extra simulation time to relax the system. In order to study the water at the active site during the process of the SAM approaching and binding until the catalytic ready stage, we used the water flooding approach [43], which has been used effectively in modeling challenging systems and does so much faster than the more rigorous grand canonical Monte Carlo (GCMC) [44]. The most likely water (below we use water which may also refer deprotonated water, namely hydroxide) configuration and the number of water molecules at the active site within 5 Å of Tyr357 phenol oxygen were examined. PTK_{Kme0} indicates the distance between the N_e of methyl lysine and the atom S of SAM as a reflection of the approaching of SAM. When there is no SAM at the active site (Fig. 3A), there are most likely eight water molecules within the 5 Å of the Tyr357 phenol oxygen. During the approaching and binding of SAM, there is at most five water molecules (Fig. 3B) left at

![Fig. 2.](image-url) The average pKa value of Tyr357 along the consistent effective dielectric constant at the Kme0 with SAM (w/ SAM) and without SAM (w/o SAM).

![Fig. 3.](image-url) The likely configuration of the active site with water. Tyr357 and methyl lysine (Kme0) are shown as stick and colored by element (C: blue; O: red; H: white). Water molecules are shown as red spheres. SAM is shown as stick and colored by element (C: yellow; O: red; N: blue).
$d_{N_{-}S} = 9$ Å and only 4 at $d_{N_{-}S} = 8$ Å. As $d_{N_{-}S}$ decreased from 8 Å to 5 Å, the water and PRDM9 went through significant re-organization during the SAM binding to reach the catalytic ready stage, which may help balance the energetic penalty associated with SAM binding.

The water flooding approach provides us with key information about the likelihood of the water (or hydroxide) configuration at the active site. There are 4 water molecules at the catalytic ready active site, two of which are close to the phenol oxygen of Tyr276 and Tyr341. We found that the two water molecules have pushed the phenol oxygen away from its crystal position and took their place. By scrutinizing the conformation, we doubt that these two water molecules may be false positive due to our long water flooding simulation and the hard wall implemented. Actually, mutation of Y276 and Y341 affects the reaction significantly, almost completely abolish it [13], which implies a proton transfer role of these two residues for the dimethylation and tri-methylation, and the two water molecules there perfectly fit the proton transfer requirement. However, further water flooding simulations done on the Kme1 and Kme2 systems indicate that the water molecules for Y276 and Y341 mostly are freshwater coming from bulk solvent after the product release and brought in during a new cofactor (SAM) approaching the active site. Additionally, the current X-ray structure (4C1Q) on the Kme1 system clearly support the Kme1...Tyr341...O proton relay network. Due to lack of the crystal structure of the active site in the Kme0 active form, and since our focus here is the first methylation step, the two water molecules will need to be studied in the future. The third water molecule close to Tyr357 is more than 3.5 Å away and excludes the direct interaction with Tyr357. Hence, for the Tyr357 proton transfer, there is a high likelihood of one water molecule at the active site, more precisely, one hydroxide in close proximity of Tyr357 (see Fig. 4). This assumption is based on the previous study of DIM-5, which shows an optimal pH 10 for its reaction with the deprotonated tyrosine to facilitate substrate lysine deprotonation [49].

This is evident, at product (Kme1 and SAH) stage, using the water flooding approach, since we found there is one water molecule close to Tyr357 at the active site bound to nitrogen of SAH. Additionally, the optimal pH is at 8.0 for SET7/9 [27] and 8.5 for PRDM9 [16], which further implies one hydroxide near Tyr357. This is also consistent with the study of the PRDM9 SAH inhibitor complex [50].

We have excluded the hydroxide abstraction of the proton directly from Kme0 based on the experimental observation that Y357F abolishes the enzyme activity. Here, we focus on the proton transfers (PT1) from Tyr357 to hydroxide, which generates deprotonated Tyr357 and water, and (PT2) from Kme0 to deprotonated Tyr357, resulting in deprotonated Kme0 and Tyr357. QM/MM calculation at the M062X/6-31+G* level indicated that the low $pK_a$ value of Tyr357 and hydroxide forms an activated H tunneling state as shown in Fig. 5A. During the approaching of SAM, the polarized hydroxide abstracts the proton to form the water molecule and hydrogen bonding to SAM resulting in deprotonated Tyr357. This deprotonated Tyr357 forms another activated H tunneling state with Kme0 (Fig. 5B). The binding rearrangement of SAM disrupts this tunneling state, and with the approaching of the methyl group from SAM, Kme0 loses the proton to Tyr357 resulting in deprotonated Kme0 and therefore preparing for the following methyl transfer (Fig. 5C). There is a good likelihood that during the product release, this water molecule is also released to the bulk solvent with the product.

**Free energy profile**

In order to understand the proton transfer better, we needed to determine the free energy profile. The free energy, calculated at the M062X/6-31+G* level, required for the PT1 and PT2 to escape the tunneling well is around 8 kcal mol$^{-1}$ and 10.66 kcal mol$^{-1}$ in water, respectively. For PT1 and PT2, the ‘depth’ of proton tunneling donor-acceptor distance is around...
2.52Å and 2.69Å; this very narrow donor-acceptor distance may account for the low activation barrier. In general, the low activation barriers energetically favors the thermally activated escape from the tunneling well, namely proton transfer, along their H-bond. The EVB simulations were done for the system of Kme0 Tyr357 OH- SAM based on the scheme depicted in Fig.1. The calculated activation barriers for the PT1 and PT2 are 5.2kcal-mol\(^{-1}\) and 5.6kcal-mol\(^{-1}\), respectively. Here, we consider the PT1 and PT2 are sequential events, and the total PT transfer energy was estimated to be 18.7kcal-mol\(^{-1}\) in water and 10.8kcal-mol\(^{-1}\) in PRDM9. As catalysis is typically encompassed by multiple temporospatial steps spanning different regions of phase space, the activation barrier was integrated to reconstruct the schematic free energy profile projections on one reaction coordinate, which is depicted in Fig.6 to give an overview of the progress of the reaction. The total reaction free energy for spontaneous proton transfer and methyl transfer was 15.9kcal-mol\(^{-1}\), which is in the range of the experimental observed value 16.6kcal-mol\(^{-1}\) [16].

**Concluding remarks**

The PRDM family has gained noticeable interest for its functions in cellular differentiation and since its dysregulation results in hematological malignancies as either tumor suppressor or promoter of oncogenic processes [4]. Experimental evidence shows that PRDMs can directly modify histones through their methyltransferase activity or indirectly through the recruitment of chromatin remodeling complexes [51,52]. The PRDMs belong to the SET domain family of histone methyltransferases; however, the molecular mechanisms have been characterized for only a few PRDMs and little is known for the rest, especially about the (a) functional role of the uncharacterized proteins as well as the redundancy within the family; (b) targets of PRDM proteins, namely the native substrates and its specificity; (c) its interaction network that affects the signaling pathways; and (d) its detailed methyltransferase mechanisms. Collectively, this information is fundamental to better understand the biological and biochemical functions of this important family of enzymes.

It is of tremendous importance to identify the endogenous PRDM substrates and catalytic essential residues as well as how the reaction takes place, thus to provide better understanding of the interaction
network as well as for designing interventions to cure diseases. Previous studies have not addressed some of the key features, such as a water channel, which has been proposed and verified [26] for SET domain lysine methyltransferases. Actually once the product (SAH) is released, there will be substantial water flooding to the cofactor binding site and the structural fluctuation of the protein naturally will relay the proton to the solvent. Nevertheless, a water channel is not necessarily required for the proton relay.

PRDM9, as a member from the PRDM family with well-characterized enzyme activities, has been used to study the comprehensive reaction profile, which will provide a reference for studying the reaction mechanism of the whole family by pinpointing potential substrates. An array of computational efforts has been used to study the PRDM9 reaction mechanism in terms of proton and methyl transfer. Our methyl transfer reaction step shares the same profile as a recent computational study of the SET7/9 [19]. The $pK_a$ value of Tyr276, Tyr341, and Tyr357 is decreased significantly to facilitate the deprotonation of the substrate Lys at Kme0, Kme1, and Kme2 states (especially when the SAM present). Our study sheds light on the surrounding tyrosines around methyl lysine, which are key to the proton transfer by bridging two tunneling ready reactant states with the binding of SAM, which triggers the cascading proton transfers. By constructing the progressive whole reaction profile, our study provides a key reference for the other member of PRDM family with undefined substrate specificity and enzyme activities.

Overall, the flexibility of the SET domain and the lack of the solution structure of the active form of the enzyme, make it a tough challenge to trace the movement of the hydrogens and hence to deduce the likely deprotonation mechanism. A recent structural study of the actin histidine methyltransferase SETD3 shows that the deprotonated Tyr312(O$^-$) and the water molecule could thereby facilitate the deprotonation of the substrate lysine [53], which is consistent with our proposed deprotonation mechanism. However, more detailed advances in structural and biochemical data are required to further address the deprotonation of the methyl lysine and the remaining questions, such as (a) how the mono-methylated lysine rotates for the next step of dimethylation; and (b) whether enzymes from the same family with uncharacterized activity have tyrosines at the active site. Meanwhile, our study provides a unique insight into the reaction mechanism of the deprotonation; however, we note that our study is based on PRDM9, and careful consideration is necessary when we apply this knowledge to other members of the methyltransferases family.

In summary, PRDM methyltransferases have been implicated in a variety of diseases. The methyltransferase activity has only been reported for 6 out of 17 members of PRDM family and the activities of the others are not known. The methyltransferase reaction consists of two steps: (i) the absolutely required lysine deprotonation and (ii) the transfer of the methyl group to the deprotonated lysine. How the methyl lysine is deprotonated has been elusive. Here, we have uncovered cascading proton transfers as a key to deprotonate the lysine residue. Our study has unravelled the unanswered question about the deprotonation of the methyl lysine in methyltransferases, which will provide a key reference in predicting the methyltransferase activity and functionality for the remaining PRDM family members.

Acknowledgement

This work was supported by an Agency for Science, Technology and Research (A*STAR) International Fellowship (AIF) awarded to LNZ. Additionally, LNZ acknowledges the computational support provided by Dr. Warshel’s laboratory at the University of
Southern California. Special thanks go to Dr. Zhen T. Chu, Dr. Chen Bai, Dr. Mikolaj Feliks, Dibyendu Mondal and Dr. Vesselin Kolev. All the authors thank Prof. Masoud Vedadi for his comments and suggestions. The helpful discussions with the members of Guccione’s laboratory are highly appreciated, a special thank you to Dr. Ernesto Guccione and Dr. Dave Keng Boon Wee for their support. In the end, the authors thank the reviewers for comments, which not merely helped to improve this work, but also will influence our future work.

Author contributions

LNZ conceived and directed the study, did the simulations and acquired the data, analyzed the data, and wrote and edited the manuscript; PK wrote and edited the manuscript.

References

1 Jenuwein T, Laible G, Dorn R and Reuter G (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci 54, 80–93.
2 Trievel RC, Beach BM, Dirk LMA, Houtz RL and Hurley JH (2002) Structure and catalytic mechanism of a SET domain protein methyltransferase. Cell 111, 91–103.
3 Dillon SC, Zhang X, Trievel RC and Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 6, 227.
4 Fog CK, Galli GG and Lund AH (2011) PRDM proteins: important players in differentiation and disease. BioEssays 34, 50–60.
5 Zhang X, Wen H and Shi X (2011) Lysine methylation: beyond histones. Acta Biochim Biophys Sin 44, 14–27.
6 Strahl BD and Allis CD (2000) The language of covalent histone modifications. Nature 403, 41–45.
7 Mosammaparast N and Shi Y (2010) Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu Rev Biochem 79, 155–179.
8 Greer EL and Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. Nature Rev Genet 13, 343–357.
9 Herz H-M, Garruss A and Shilatifard A (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem Sci 38, 621–639.
10 Rizzo PAD and Trievel RC (2011) Substrate and product specificities of SET domain methyltransferases. Epigenetics 6, 1059–1067.
11 Hohenauer T and Moore AW (2012) The Prdm family: expanding roles in stem cells and development. Development 139, 2267–2282.
12 Guccione E and Richard S (2019) The regulation, functions and clinical relevance of arginine methylation. Nat Rev Mol Cell Biol 20, 642–657. https://doi.org/10.1038/s41580-019-0155-x
13 Wu H, Mathioudakis N, Diourgoura B, Dong A, Dombrovskii L, Baudat F, Cusack S, deMassy B and Kadlec J (2013) Molecular basis for the regulation of the H3K4 Methytransferase activity of PRDM9. Cell Rep 5, 13–20.
14 Blazer LL, Lima-Fernandes E, Gibson E, Eram MS, Lopprau P, Arrowsmith CH, Schapira M and Vedadi M (2016) PR Domain-containing Protein 7 (PRDM7) is a histone 3 lysine 4 trimethyltransferase. J Biol Chem 291, 13509–13519.
15 Mzoughi S et al. (2017) PRDM15 safeguards naive pluripotency by transcriptionally regulating WNT and MAPK-ERK signaling. Nat Genet 49, 1354–1363.
16 Eram MS, Bustos SP, Lima-Fernandes E, Sarayheva A, Senisterra G, Hajian T, Chau I, Duan S, Wu H, Dombrovski L et al. (2014) Trimethylation of Histone H3 Lysine 36 by Human Methyltransferase PRDM9 Protein. J Biol Chem 289, 12177.
17 Kwon T, Chang JH, Kwak E, Lee CW, Joachimiak A, Kim YC, Lee J and Cho Y (2003) Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. EMBO J 22, 292–303.
18 O’Hagan D and Schmidbergera JW (2010) Enzymes that catalyse SN2 reaction mechanisms. Nat Product Rep 27, 900–918.
19 Guo H-B and Guo H (2007) Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. Proc Natl Acad Sci USA 104, 8797–8802.
20 Grimsley GR, Scholtz JM and Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 18, 247–251.
21 Xiao B, Wilson JR and Gamblin SJ (2003) SET domains and histone methylation. Curr Opin Struct Biol 13, 699–705.
22 Hu P and Zhang Y (2006) Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: an ab initio QM/MM–FE study with multiple initial structures. J Am Chem Soc 128, 1272–1278.
23 Wilson JR, Jing C, Walker PA, Martin SR, Howell SA, Blackburn GM, Gamblin SJ and Xiao B (2002) Crystal structure and functional analysis of the histone methyltransferase SET7/9. Cell 111, 105–115.
24 Horowitz S, Dirk LM, Yesselman JD, Nimtz JS, Adhikari U, Mehl RA, Scheiner S, Houtz RL, AlHashimi HM and Trievel RC (2013) Conservation
Catalytic mechanism of SAM-dependent methyltransferases

L. N. Zhao and P. Kaldis

and functional importance of carbon-oxygen hydrogen bonding in AdoMet-dependent methyltransferases. J Am Chem Soc 135, 15536–15548.

Chu Y, Sun L and Zhong S (2015) How Y357F, Y276F mutants affect the methylation activity of PRDM9: QM/MM MD and free energy simulations. J. Mol. Model 21, 125.

Zhang X, Bruice TC (2007) Histone lysine methyltransferase SET7/9: formation of a water channel precceeds each methyl transfer. Biochemistry 46, 14834–14844.

Zhang X and Burice TC (2008) Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. Proc Natl Acad Sci USA 105, 5728–5732.

OpenEye Toolkits (2019) OpenEye Scientific Software. https://www.eyesopen.com/toolkit-development

Warshel A and Weiss RM (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. J Am Chem Soc 102, 6218–6226.

Kamerlin SCL and Warshel A (2010) The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. Faraday Discuss 145, 71–106.

Zhao LN, Mondal D and Warshel A (2020) Exploring alternative catalytic mechanisms of the Cas9 HNH domain. Proteins 88, 260–264.

Aqvist J and Warshel A (1989) Calculations of free energy profiles for the Staphylococcal nuclease catalyzed reaction. Biochemistry 28, 4680–4689.

Lee FS, Chu ZT and Warshel A (1993) Microscopic and semi-microscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. J Comput Chem 14, 161–185.

Lee FS and Warshel A (1992) A local-reaction-field method for fast evaluation of long-range electrostatic interactions in molecular simulations. J Chem Phys 97, 3100–3107.

King G and Warshel A (1990) Investigation of the free energy functions for electron transfer reactions. J Chem Phys 93, 8682.

Warshel A (1997) Computer Modeling of Chemical Reactions in Enzymes and Solutions. New York, NY: John Wiley & Sons, Inc. https://www.wiley.com/en-us/Computer+Modeling+of+Chemical+Reactions+in+Enzymes+and+Solutions-p-9780471184409

Singh N and Warshel A (2010) Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. Proteins Struct Funct Bioinf 78, 1705–1723.

Schutz CN and Warshel A (2001) What are the dielectric “constants” of proteins and how to validate electrostatic models? Proteins: Struct Funct Bioinf 44, 400–417.

Sham YY, Chu ZT and Warshel A (1997) Consistent calculations of pKa’s of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B 101, 4458–4472.

Warshel A, Sharma K, Kato M and Parson WW (2006) Modeling electrostatic effects in proteins. BBA – General Subjects 1764, 1647–1676.

Chakrabarty S and Warshel A (2013) Capturing the energetics of water insertion in biological systems: The water flooding approach. Proteins Struct Funct Bioinf 81, 93–106.

Yoon H, Kolev V and Warshel A (2017) Validating the water flooding approach by comparing it to grand canonical Monte Carlo simulations. J Phys Chem B 121, 9358–9365.

Frisch MJ et al. (2013) Gaussian, Inc. Wallingford CT. https://gaussian.com/g09citation/

Holliday GL, Mitchell JBO and Thornton JM (2009) Understanding the functional roles of amino acid residues in enzyme catalysis. J Mol Biol 390, 560–577.

Yuan S, Roy KL, Venken T, Lammens W, denEnde WV and Maeyer MD (2012) pKa modulation of the acid/base catalyst within GH32 and GH68: a role in substrate/inhibitor specificity? PLoS ONE 7, e37453.

Yu H and Griffiths TM (2014) pKa cycling of the general acid/base in glycose hydrolase families 33 and 34. Phys Chem Chem Phys 16, 5785.

Zheng X, Tamaru H, Khan S, Horton J, Keefe L, Selker E and Cheng X (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell 111, 117–127.

Allali-Hassani A, Szewczyk MM, Ivanochko D, Organ SL, Bok J, Ho JSY, Gay FPH, Li F, Blazer L, Eram MS et al. (2019) Discovery of a potent, selective and cell active PRDM9 inhibitor. Nat Commun 10, 5759. https://doi.org/10.1038/s41467-019-13652-x

Ivanochko D, Halabelian L, Henderson E, Savitsky P, Jain H, Marcon E, Duan S, Hutchinson A, Seiota A, Barsyte-Lovejoy D et al. (2019) Direct interaction between the PRDM3 and PRDM16 tumor suppressors and the NuRD chromatin remodeling complex. Nucleic Acids Res 47, 1225–1238.

Zazoo ED, Rosa CD, Abbondanza C and Monchamont B (2013) PRDM proteins: molecular mechanisms in signal transduction and transcriptional regulation. Biology 2, 107–141.

Dai S, Horton JR, Woodcock CB, Wilkinson AW, Zhang X, Gozani O and Cheng X (2019) Structural basis for the target specificity of actin histidine methyltransferase SETD3. Nat Commun 10, 3541.
Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. The RMSD of the MD/EVB simulation.

Fig. S2. The relative free energy (kcal/mol) calculated for the methyl transfer to Kme0, Kme1 and Kme2 in water.

Fig. S3. The EVB resonance states ($\Phi^1_{Kme0}$, $\Phi^2_{Kme0}$, $\Phi^3_{Kme0}$, $\Phi^4_{Kme1}$, $\Phi^5_{Kme1}$, $\Phi^6_{Kme1}$, $\Phi^7_{Kme2}$ and $\Phi^8_{Kme2}$) used to describe the methyl transfer catalytic reaction.

Fig. S4. The average pKa value of Tyr276 along the consistent effective dielectric constant at the Kme0, Kme1 and Kme2 with SAM (w/SAM) and without SAM (w/o SAM).

Fig. S5. The average pKa value of Tyr341 along the consistent effective dielectric constant at the Kme0, Kme1 and Kme2 with SAM (w/SAM) and without SAM (w/o SAM).

Fig. S6. The average pKa value of Tyr357 along the consistent effective dielectric constant at the Kme0, Kme1 and Kme2 with SAM (w/ SAM) and without SAM (w/o SAM).

Fig. S7. The average pKa value of Tyr357 along the consistent effective dielectric constant at the Kme0 with SAM (w/ SAM), with SAH (w/ SAH) and without SAM (w/o SAM).

Fig. S8. The EVB atoms used to define different resonance states for methyl transfer.

Table S1. Calculated apparent pKa value of the three tyrosine residues in the WT PRDM9 at $e_{eff} = 8$.

Table S2. The protonation state (in terms of charge) of key ionizable residues.

Table S3. Partial charges used in the methyl transfer mechanism.

Table S4. EVB parameters: Part 1.

Table S5. EVB parameters: Part 2.

Table S6. EVB parameters: Part 3.

Table S7. EVB parameters: Part 4.

Table S8. EVB parameters: Part 5.