Supporting Information

# Exploring the origin of the catalytic power and product specificity of SET domain protein methyltransferase

*A. H. Lima,[a] C. N. Alves,[a] R. Prasad,[b*] and J. Lameira[a*]*

Faculdade de Biotecnologia e Laboratório de Planejamento e Desenvolvimento de Fármacos; Universidade Federal do Pará, 66075-110, Belém, PA, Brazil.

Southern California, Department of Chemistry, SGM 418, 3620 McClintosk Avenue, Los Angeles, California 90089, United States

## S1: Methods

### The EVB method

In order to compute the free energy differences and activation barriers of the first and second methyl transfer for native SET8 and its mutant Y334F, we have used empirical valence bond (EVB) method[1]. The EVB method provides a powerful way of obtaining reliable activation free energies, while taking into account the full protein flexibility and configuration space. In this method, the reaction is described by diabatic states that correspond to classical valence bonds (VB) structures which represent the reactants (or intermediate) and product states, where the reaction free energy profile is calculated on the ground state energy surface $E_g$. In the case of our particular reaction, where we have a $S_N2$ reaction we have two diabatic states for each round of methylation:

The potential energy functions of the diabatic valence bond states and their mixing term are represented by:

$$H_{ii} = \epsilon_i = \alpha_{gas}^{\ i} + U_{intra}^{\ i}(R,Q) + U_{Ss}(R,Q,r,q) + U_{ss}(r,q) \quad (1)$$

The off-diagonal elements are usually represented by:

$$H_{ij} = A_{ij} e^{[-\mu_{ij}(r_{ab} - r_{ab,0})]} \quad (2)$$

Where $A_{ij}$, $\mu_{ij}$ and $r_{ab,0}$ are empirical constants and its values are calibrated on the basis of the computational reproduction of the experimental free energy profile. In this equation, $r_{ab}$ represents the distance between the two atoms characterizing the affected bond the *i*th and *j*th states. In addition, the $H_{ij}$ elements are assumed to be the same in the gas phase, solution and protein. The adiabatic ground-state energy $E_g$ and the corresponding eigenvector $C_g$ are obtained by solving the secular equation:

$$H_{EVB}C_g = E_gC_g \quad (3)$$

The EVB treatment provides a natural picture of intersecting electronic states that is useful for exploring the effect of environment on the chemical reaction in condensed phases. In the EVB approach the system is changed from reactant to products adiabatically mapping it from one diabatic state to another. That is, In order to simulate the formation of the chemical bond

during the transition between two EVB states, $\varepsilon_1$ and $\varepsilon_2$ (initial and final states), we conduct MD simulations of the system on a mapping potential, $\varepsilon_m$, that is determined by a linear combination of the initial and final states:

$$\varepsilon_m = (1-\lambda_m)\varepsilon_1 + \lambda_m\varepsilon_2 (0 \le \lambda_m \le 1) \tag{4}$$

Where $\lambda_m$ is an order parameter going from 0 to 1 in N+1 windows as the initial state is changed to the final state. The free energy change between the consecutive steps can be calculated by the FEP procedure as [2]:

$$\Delta G_{m\to m+1} = -\beta^{-1} ln\left\langle e^{\left\{-\beta\left[\varepsilon_m(\lambda_{m+1})-\varepsilon_m(\lambda_m)\right]\right\}}\right\rangle_m \tag{5}$$

After the completion of MD-FEP calculations, the free energy functional that corresponds to the adiabatic ground state surface $E_g$ is then obtained by FEP-umbrella sampling (FEP/US) method[2,3] that can be written as:

$$\Delta g(x') = \left\langle \sum_{m=0}^{i-1} \Delta G_{m\to m+1} - \beta^{-1} ln\left\langle \delta(x-x') e^{\left\{-\beta\left[E_g-\varepsilon_m(\lambda_i)\right]\right\}}\right\rangle_{\varepsilon_m}\right\rangle_i \tag{6}$$

Where $\sum_{m=0}^{i-1} \Delta G_{m\to m+1}$ represents the free energy difference between the first and *i*th mapping potential, $\delta$ correspond to Dirac's delta function. $\varepsilon_m$ is the mapping potential that keeps $x$ in the region of $x'$. The generated reaction coordinate, x, is usually taken as the energy gap ($x = \varepsilon 1 - \varepsilon 2$). This selection is particularly powerful when on tries to represent all of the many dimensional solvent space by a single coordinate. Additionally, it is important to note that the diabatic free energy profiles of the reactant and product represent microscopic equivalent [4] of the Marcus parabolas. [5]

**The LRA Method**

The linear response approximation (LRA) treatment[6,7] provides a good estimation for the free energy associated with the change between two potential surfaces ($U_1$ and $U_2$) by

$$\Delta G(U_1 \to U_2) = \frac{1}{2}\left(\langle U_2 - U_1\rangle_1 + \langle U_2 - U_1\rangle_2\right) \tag{S7}$$

The notation $\langle\rangle_i$ designates an average over trajectories propagated on the potential energy surface $U_i$. Herein, we have used the LRA approach to calculate the free energy of changing the charges of a given state from actual EVB charges (Q) to the charges of nonpolar state (zero residual changes).

**The system**

The initial structure for calculations were taken from Protein Data Bank (PDB) with from 3F9X[8] PDB code. The simulation systems were solvated using the surface constrained all atom solvent ~~model~~ (SCAAS) model, using a water sphere with a radius of 20 Å centered on the substrate and surrounded first by a 2 Å grid of Langevin dipoles, and then by a bulk solvent. Further information about SCAAS model can be found in Ref. 11 and 12 of this material. It is also important to clarify that for the first methylation, we have 4054 atoms in the system, where 17 atoms are in the region I. For the second methylation we have 4009 atoms in the system, where 20 atoms are in the region I. The free energy perturbation mapping was performed in 41 frames of 20 ps length each for the movement along the reaction coordinate, using the SCAAS model, after the respective system underwent a 1500 ps relaxation run. Here, the free energy was obtained using the canonical (NVT) ensemble which can be generalized to the Gibbs free energy (NPT) and other thermodynamic ensembles. All simulations were performed at 300 K using a 1 fs time step. In order to obtain reliable sampling, the simulations were repeated at least five times with different initial conditions (obtained from arbitrary points in the relaxation trajectory after the initial 1500 ps relaxation run) for each reacting system. Weak residue constraints of 0.3 kcal $mol^{-1}$ $Å^{-2}$ were applied to region I that involves part of SAM and Lys20. All EVB calculations were carried out by the MOLARIS simulation program[9,10] using the ENZYMIX force field. It is worth mentioning that we have selected ionized residues within the first solvation shells of the SAM, including Tyr245, which presented pKa of 9.4 in accordance with PDLD/S-LRA pKa calculations. This ionized residue forms a stable salt bridge with Lys20. Indeed, our free energies results suggest that the activation barrier for methyl transfer form SAM to Lys22 is unfavorable when Tyr245 is protonated. In addition, the tautomers of histidine residues (in its neutral form) were determined automatically by our standard procedure (included in MOLARIS), which selects automatically the configuration with lower electrostatic energy. Finally, Long-range electrostatic effects were treated by the local reaction field (LRF) method.[9]

**REFERENCES:**


(1) Warshel, A.; Weiss, R. M. Journal of the American Chemical Society1980, 102, 6218.

(2) Warshel, A. Computer modeling of chemical reactions in enzymes and solutions New York, 1991.

(3) Aqvist, J.; Warshel, A. Chemical Reviews1993, 93, 2523.

(4) King, G.; Warshel, A. Journal of Chemical Physics1990, 93, 8682.

(5) Marcus, R. A. Journal of Chemical Physics1965, 43, 679.

(6) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. Protein Engineering1992, 5, 215.

(7) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Biochimica Et Biophysica Acta-Proteins and Proteomics2006, 1764, 1647.

(8) Couture, J.-F.; Dirk, L. M. A.; Brunzelle, J. S.; Houtz, R. L.; Trievel, R. C. Proceedings of the National Academy of Sciences of the United States of America2008, 105, 20659.

(9) Lee, F. S.; Chu, Z. T.; Warshel, A., Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *Journal of Computational Chemistry* **1993,** *14* (2), 161-185

(10) Warshel, A. C. C., Z. T.; Villa, J.; Strajbl, M.; Schutz, C. N.; Shurki, A. V., S.; Chakrabarty, S.; Plotnikov, N. V.; Schopf, P; 9.11 ed.; University of Southern California: Los Angeles, CA, 2012.

(11) G. King and A. Warshel, “A surface constrained all-atom solvent model for effective simulations of polar solutions,” J. Chem. Phys.,vol. 91, no.6, pp.3647–3661, 1989.

(12) F.S. Lee, Z.T. Chu, and A. Warshel, “Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the polaris and enzymix programs,” J. Comput. Chem.,vol. 14, pp.161–185, 1993.

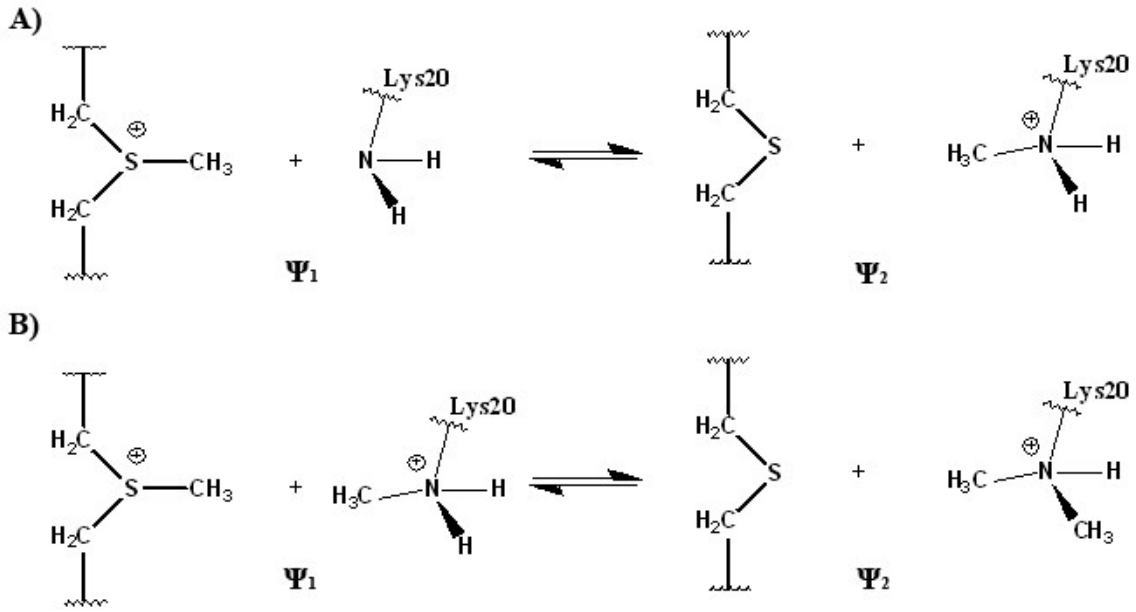


**Fig. S1.** The valence bond structures representing the reactants and products used for (A) First and (B) second methyl transfer.

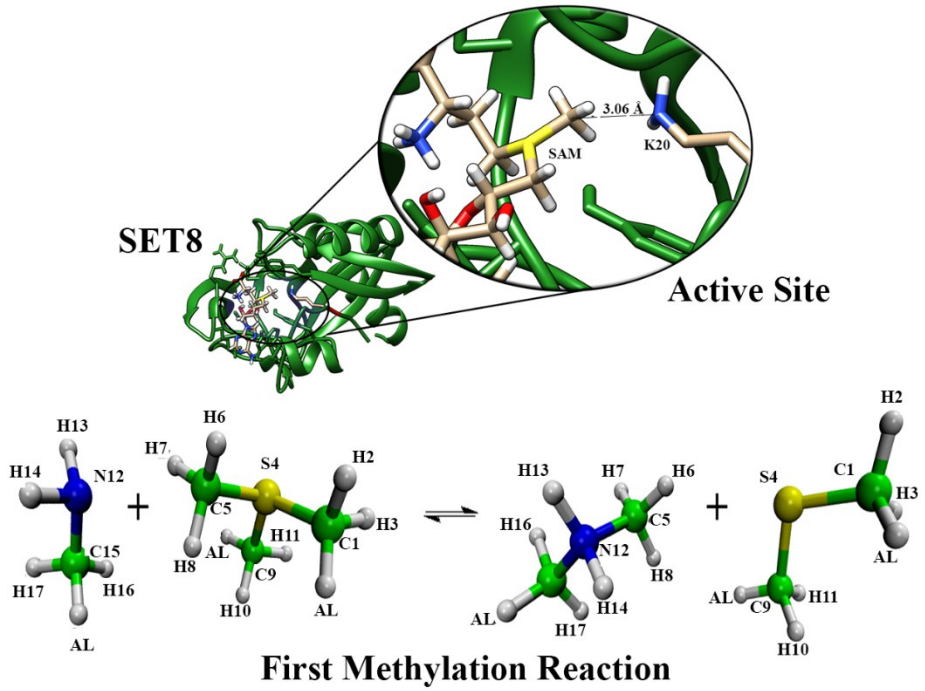


**Fig. S2.** Active site of the SET8 enzyme and its reaction for the first methylation.

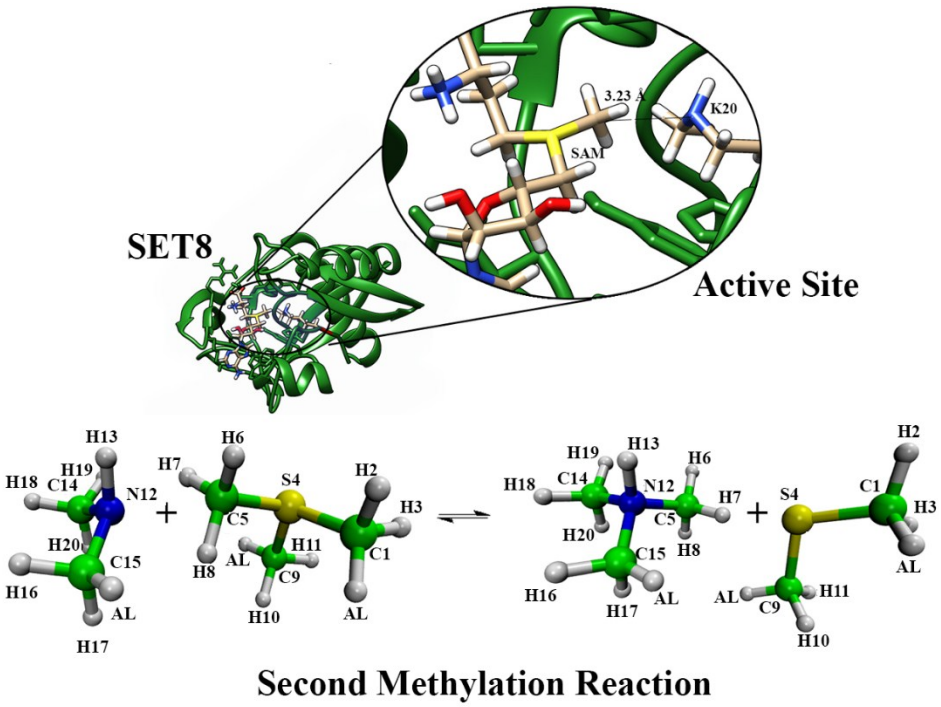


**Fig. S3.** Active site of the SET8 enzyme and its reaction for the second methylation.

**Table S1.** Activation free energies computed with B3LYP/6-31G(d,p) and MP2/6-31G(d,p) levels of theory in gas-phase and water for first and second methylation, where the solvation effects were computed with single-point PCM calculations and the thermal corrections were obtained from HF calculations using the same basis set.

| Reaction | MP2/6-31+G(d,p) gas | B3LYP/6-31+g(d,p) gas | MP2/6-31+g(d,p) PCM | B3LYP/6-31+g(d,p) PCM |
|---|---|---|---|---|
| First Methylation | | | | |
| RC | 0 | 0 | 0 | 0 |
| TS | 20.78 | 13.84 | 22.49 | 17.64 |
| PC | -10.52 | -31.751 | -24.63 | -20.41 |
| Second Methylation | | | | |
| RC | 0 | 0 | 0 | 0 |
| TS | 18.61 | 12.72 | 18.44 | 18.50 |
| PC | -15.15 | -18.95 | -18.77 | -21.832 |

**Table S2.** Charges used in first methylation.[(a)]

| | **Reactant State** | | **Product State** | |
|---|---|---|---|---|
| Atom number | AtomType | Charge | AtomType | Charge |
| 1 | C0 | -0.266655 | C0 | -0.195741 |
| 2 | H0 | 0.223713 | H0 | 0.167843 |
| 3 | H0 | 0.223713 | H0 | 0.167843 |
| 4 | S+ | 0.415429 | S0 | -0.213389 |
| 5 | C0 | -0.490368 | C0 | -0.308306 |
| 6 | H0 | 0.223713 | H0 | 0.159612 |
| 7 | H0 | 0.223713 | H0 | 0.159612 |
| 8 | H0 | 0.223713 | H0 | 0.159612 |
| 9 | C0 | -0.266655 | C0 | -0.195741 |
| 10 | H0 | 0.223713 | H0 | 0.167843 |
| 11 | H0 | 0.223713 | H0 | 0.167843 |
| 12 | N0 | -0.800262 | N1 | -0.058837 |
| 13 | H0 | 0.323809 | H0 | 0.325638 |
| 14 | H0 | 0.323809 | H0 | 0.325638 |
| 15 | C0 | 0.188069 | C0 | -0.148694 |
| 16 | H0 | 0.003415 | H0 | 0.159612 |
| 17 | H0 | 0.003415 | H0 | 0.159612 |

[(a)] Charges in au.

**Table S3.** Charges used in second methylation.[(a)]

| | Reactant State | | Product State | |
|---|---|---|---|---|
| Atom number | AtomType | Charge | AtomType | Charge |
| 1 | C0 | -0.296365 | C0 | -0.379542 |
| 2 | H0 | 0.235763 | H0 | 0.228063 |
| 3 | H0 | 0.235763 | H0 | 0.228063 |
| 4 | S+ | 0.433221 | S0 | -0.113180 |
| 5 | C0 | -0.532128 | C0 | -0.272555 |
| 6 | H0 | 0.235763 | H0 | 0.145749 |
| 7 | H0 | 0.235763 | H0 | 0.145749 |
| 8 | H0 | 0.235763 | H0 | 0.145749 |
| 9 | C0 | -0.296365 | C0 | -0.379542 |
| 10 | H0 | 0.235763 | H0 | 0.228063 |
| 11 | H0 | 0.235763 | H0 | 0.228063 |
| 12 | N0 | -0.477417 | N1 | 0.135540 |
| 13 | H0 | 0.326556 | H0 | 0.330398 |
| 14 | C0 | -0.065914 | C0 | -0.126806 |
| 15 | C0 | -0.146910 | C0 | -0.272555 |
| 16 | H0 | 0.080996 | H0 | 0.145749 |
| 17 | H0 | 0.080996 | H0 | 0.145749 |
| 18 | H0 | 0.080996 | H0 | 0.145749 |
| 19 | H0 | 0.080996 | H0 | 0.145749 |
| 20 | H0 | 0.080996 | H0 | 0.145749 |

[(a)] Charges in au.

**Table S4.** Bond parameters: $\Delta M = D[1 - e^{\{-a(b-b_0)\}}]^2$ [(a)]

| Bond Type | $D$ | $b_0$ | $a$ |
|---|---|---|---|
| C0-S0 | 70.0 | 1.80 | 2.0 |
| C0-S1 | 70.0 | 1.80 | 2.0 |
| C0-N0 | 93.0 | 1.44 | 2.0 |
| C0-N1 | 100.0 | 1.48 | 1.0 |

[(a)] The parameter $D$ is the Morse potential, $b_0$ is the distance and the parameter "$a$" is calibrated by adjusting its value to reproduce the observed vibrational frequency of given bond. Energies in kcal/mol, distance in Å.

**Table S5.** Bond Angles parameters: $U_0 = \frac{1}{2}K_0(\theta - \theta_0)^2$ (a)

| Bond Type | $\theta$ | $K_\theta$ |
|---|---|---|
| X-S0-Y | 102.5 | 100.0 |
| X-S1-Y | 103.5 | 100.0 |
| X-C0-Y | 109.5 | 50.0 |
| X-N1-Y | 109.5 | 50.0 |
| X-N0-Y | 109.5 | 50.0 |

Energies in kcal/mol, angles in degree.

**Table S6.** Nobonded parameters: $U_{nb} = A_i A_j r^{-12} - B_i B_j r^{-6}$ (a)

| Bond Type | $A$ | $B$ |
|---|---|---|
| H0 | 7.00 | 0.00 |
| S0 | 1831.8 | 23.0 |
| S1 | 1022.5 | 23.0 |
| C0 | 632.0 | 24.0 |
| N1 | 774.0 | 24.0 |
| N0 | 774.0 | 24.0 |

(a) The parameter A represents the hard core repulsion and B represent s the van der Waals attraction. The $U_{nb}$ is used for describing interaction between solute and the solvent.

**Table S7**. Nobonded parameters: $U_{nb} = Ae^{\{-ar\}}$ (a)

| Bond Type | $A$ | $a$ |
|---|---|---|
| C0…S0 | 19999.0 | 4.0 |
| C0…S1 | 19999.0 | 4.0 |

(a) The parameters are used for describing interaction between solute atoms.

**Table S8.** EVB off-diagonal and shift parameters.(a)

| **EVB off-diagonal and shift parameters** | |
|---|---|
| First Methylation | Second Methylation |
| A=-31.5 | A=-34.5 |
| α= -83.0 | α= -40.0 |
| μ=0.0 | μ=0.0 |

(a) where, the α is the gas phase shift and $H_{12}$ is the matrix element between the two states