Molecular Mechanism for Eliminylation, a Newly Discovered Post-Translational Modification

Zhihong Ke,[1] Gregory K. Smith,[1] Yingkai Zhang,[2] and Hua Guo[1,]*

[1]Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131

[2]Department of Chemistry, New York University, New York, New York, 10003

Supporting Information

Table of Content:

*: corresponding author: hguo@unm.edu

I. Method

The initial structure of the SpvC enzyme-substrate (ES) complex was constructed based on the X-ray structure 2Q8Y, which is a K136A mutant co-crystallized with dual-phosphorylated peptide substrate.[1] As in our earlier work,[2] the unresolved residues 1 - 26 of the enzyme were removed, while the absent residues Ser96 and Gln97 were rebuilt from another crystal structure (PDB code: 2Z8P).[3] Ala136 was mutated back to neutral Lys136 with the mutagenesis function in PyMOL (www.pymol.org). Hydrogen atoms were added by Leap in Amber.[4] By carefully checking the local hydrogen bond network, the histidine residues in SpvC were determined to be protonated as: HID32, HID41, HID51, HID81, HIE88, HIP106, HIP174, HIP203. Our choice of the protonation states for Lys136 and His106 was based on those proposed based on experimental pH profiles of the SpvC catalysis.[1] The model of the Y158F mutant was prepared using the same protocol, except the hydroxyl group in Tyr158 is replaced with a hydrogen atom.

The ES complex was then solvated in a periodic rectangular water box of 67 × 69 × 66 $Å^3$, with a buffer distance of 10 Å between the wall and the closest protein atom in each direction, which led to a system of 30107 atoms in total. The protein charges were neutralized with three Na^+ ions. The added solvent molecules and ions were first minimized and then equilibrated by 50 ps NVT and 50 ps NPT molecular dynamics (MD) simulations with the heavy atoms of protein and crystal water restrained by a force constant of 50 $kcal \cdot mol^{-1} \cdot Å^{-2}$. Then the protein was optimized by releasing the restraint gradually. Finally, a 2.0 ns non-restrained MD simulation was carried out. Throughout the classic MD simulation, the AMBER99SB force field[5-6] and TIP3P

model[7] for water molecules were employed with a time step of 1 fs. Force field parameters and the charges of the phosphotyrosine and phosphothreonine were adapted from Ref. 8. An 8 Å cutoff was introduced for nonbonding interactions, and the particle mesh Ewald (PME) method[9-10] was employed to treat long-range electrostatic interactions. The SHAKE algorithm[11] was applied to constrain all covalent bonds involving hydrogen. All classical calculations were performed with Amber 10.[4]

The QM/MM model was prepared from a MD snapshot by removing water molecules outside of a 27 Å radius from the C_α atom of the phosphothreonine. The resulting system of 8882 atoms was subjected to spherical boundary conditions, in which only the atoms within a 20 Å radius of the C_α atom of the phosphothreonine were allowed to move. The QM region (colored blue in Fig. 1) includes phosphothreonine and the side chains of His106 and Lys136, totaling 52 atoms. Two neighboring peptide bonds are also included. The QM region was treated at the B3LYP/6-31G(d) level of theory, and the inclusion of diffuse functions does not change the mechanism. All the other atoms in the MM subsystem were described with the same force field employed in the classical MD. The QM-MM boundaries were treated with the pseudo-bond approach.[12-15] Specifically, the $C(sp^3)$-$C(sp^2)$ and $C(sp^3)$-$N(sp^3)$ pseudo-bond parameters have been employed for the backbone boundaries, respectively. The total number of basis functions for the wild type SpvE model is 416. A cutoff of 12 Å was employed for van der Waals interactions in the MM subsystem, while a cutoff of 18 Å was for electrostatic interactions. No cutoff was used for electrostatic interactions between QM and MM atoms.

The catalyzed reaction was first investigated using the reaction coordinate driving (RCD) method.[13] The reaction path was obtained by driving the system along the reaction coordinate back and forth several times. Several reaction coordinates were tested and a two dimensional minimum energy surface was determined along the α proton abstraction coordinate and the C_β-O_γ bond breaking coordinate. The configurations along the reaction path were later used to calculate the potential of mean force (PMF), which includes the fluctuation of both QM and MM atoms. To this end, the reaction coordinate, which is defined as RC = $d_{C\beta\text{-}O\gamma}$ - $d_{H\alpha\text{-}N\zeta}$, was divided into 22 windows for umbrella sampling,[16] with each window biased by a harmonic potential with a force constant of 40 - 350 kcal·mol^{-1}·Å^{-2}. In the MD simulation, the MM subsystem was first equilibrated by a 500 ps MD simulation with the QM region fixed. This is followed by QM/MM MD where the energy of the entire QM/MM system was calculated on-the-fly at each time step (1 fs). For the wild-type, the QM/MM MD was run for 30 ps for each window. For the Y158F mutant, 30 ps sampling was collected in 10 windows around the transition states, while 10 ps in the other 12 windows in the reactant and product regions. In the MD, Newton's equation of motion was integrated with the Beeman algorithm.[17] The Berendsen thermostat method[18] was used to maintain the system temperature at 300 K. The PMF was determined using the weighted histogram analysis method (WHAM).[19] All the QM/MM calculations have been conducted with modified versions of QChem[20] and TINKER programs.[21]

To understand how the enzyme environment facilitates the catalysis, the electrostatic and van der Waals interactions between QM and MM subsystems were calculated at each stationary point along the reaction. For each structure, roughly 1000 snapshots were used. The

ESP charges of the QM atoms were used in computing the electrostatic interaction between the QM region and an MM residue. Furthermore, the contribution of individual residue to the catalysis was estimated by: $\Delta E_i = (E_{elec+vdw})_i^{TS} - (E_{elec+vdw})_i^{ES}$, where the negative ΔE_i indicates that the residue *i* reduces the barrier.

II. Minimal Energy Surface and Characters of Reaction Path

To address the mechanistic questions on the catalyzed reaction, we first calculated the two-dimensional minimal energy surface along the α proton abstraction coordinate and the C_β-O_γ bond breaking coordinate. The resulting surface, as shown in Fig. S1, strongly suggests that the proton abstraction occurs prior to the C_β-O_γ bond cleavage. Neither the carbocation pathway nor the concerted mechanism is viable due to high energies, though it should be stressed that this potential energy surface is not expected to be quantitatively accurate because the protein was not allowed to fluctuate. The reaction path obtained by using the reaction coordinate: $RC = d_{C\beta\text{-}O\gamma} - d_{H\alpha\text{-}N\zeta}$, also traced in Fig. S1, indicates that this RC is capable of following the minimal energy pathway. Hence, we computed the PMFs for the wild-type of SpvC and its Y158F mutant using this reaction coordinate.

The optimized structure of the enzyme-substrate (ES) complex is provided in Fig. S2 and the key geometric parameters are listed in Table S1. During the *ab initio* QM/MM MD simulation, the overall hydrogen bond network in the active site is well conserved, mimicking the X-ray structure. Specifically, the phosphoryl group of pThr is tightly held by hydrogen bonds with the positively charged cavity comprised of Lys104, Arg148, Arg213 and Arg220. His106 is well anchored by Asp201 and the phosphothreonine. Lys136 is positioned right below the C_α

atom, ready to abstract the α proton, as evidenced by the almost in-line angle C_α-H_α-N_ζ. In addition, the backbone carbonyl oxygen adjacent to C_α forms hydrogen bonds with Lys104 and Tyr158, as evidenced by the hydrogen bond distances 1.95±0.19 Å and 1.81±1.16 Å. These interactions appear to polarize the carbonyl group, as evidenced by an elongated C-O distance (1.24 ± 0.05 Å) and a shortened C-C_α distance (1.53 ± 0.07 Å), and perhaps more importantly, an elongated H_α-C_α distance (1.12 ± 0.05 Å).

The structures of the two transition states are displayed in Fig. 3, and the important internuclear distances are also listed in Table S1. TS-I involves largely the elongation of the H_α-C_α distance and the shortening of the H_α-N_ζ distance, which change from 1.12 ± 0.05 Å and 2.09 ± 0.09 Å at ES to 1.62 ± 0.16 Å and 1.21 ± 0.07 Å at TS-I, respectively. On the other hand, TS-II represents mainly the cleavage of the C_β-O_γ bond, as evidenced by the significant change on the C_β-O_γ distance from 1.51 ± 0.06 Å at INT to 1.84 ± 0.06 Å at TS-II. The INT complex, also shown in Fig. S2, is a metastable complex formed after the proton abstraction and before the C_β-O_γ bond cleavage. Interestingly, the proton transfer from His106 to the phosphate leaving group is delayed at TS-II, as evidenced by the minimal change of the H_ε-N_ε bond length from 1.03 ± 0.04 Å at ES to 1.06 ± 0.05 Å at TS-II. Only after the bond C_β-O_γ is completely broken, did the proton transfer. This is consistent with the consensus that phosphate is a reasonably good leaving group.

The enzyme-product (EP) complex, also shown in Fig. S2, features a β-methyldehydroalanine and the phosphate leaving group. The C_α-C_β distance of 1.34 ± 0.06 Å represents a typical C-C double bond. On the other hand, the C_β-O_γ distance is 3.28 ± 0.27 Å,

indicative of a completely broken bond. The phosphate dianion is firmly locked by the positively charged residues: Lys104, Arg148, Arg213 and Arg220. It also has a very strong hydrogen bond with His106, as evidenced by the bond lengths of H_ε-O_γ and H_ε-N_ε being 1.12 ± 0.18 Å and 1.53 ± 0.25 Å, respectively. The backbone carbonyl oxygen maintains hydrogen bonding with both Lys104 and Tyr158. Finally, the backbone bonds C_α-C and C_α-N become shorter, consistent with a weak conjugated system between the backbone and the bond C_α=C_β.

III. Role played by individual residues

The Lys104 residue plays a dual role, acting as part of the oxyanion hole responsible for the acidic α proton, and it also interacts with the phosphate via a hydrogen bond (Fig. 1), providing electrostatic stabilization of the leaving group. Hence, its role in the catalysis is not as clear cut as that of Tyr158. To gauge the contribution of this and other surrounding residues, we have carried out a perturbation analysis. In Fig. S3, the contributions of all residues in the MM region are displayed for both TS-I and TS-II. Recall that negative contributions lower the barrier while positive ones raise the barrier.

This analysis identified several residues that contribute favorably to the catalysis. They include positively charged residues, Lys104, Arg148, Arg213, and Arg220, which help to stabilize the phosphoryl group. It is interesting to note that their stabilization effect is much smaller for TS-I, as the proton abstraction does not change significantly the charge of the phosphoryl group. On the other hand, the cleavage of the C_β-O_γ bond at TS-II creates an additional charge in the phosphoryl leaving group, which requires stabilization by these cationic residues, resulting in a large negative ΔE_i. These findings are consistent with the mutagenesis results,

which showed that SpvC loses most or all the catalytic activity if any of these residues is replaced.[1] On the other hand, Asp84 interacts with the general base Lys136 through a water bridge and is presumably responsible for the stabilization of the cationic Lys136 after its abstraction of the proton. Finally, the impact of Tyr158 on catalysis is also consistent with its mechanistic role as an oxyanion hole.

There are only a few residues that increase the barriers, but their impacts are typically small. Lys134, for example, is known to plays a structural role in substrate recognition, forming a hydrogen bond with the substrate phosphotyrosine.[1] On the other hand, Asp201 is essential in anchoring His106.[1] Thus, their destabilizing influence on the transition states are offset by their essential role in active site preorganization. Finally, Lys142 is 14 Å from the substrate with no discernable functional role. Its apparent destabilizing influence on the transition state is likely a flucuation artifact since this residue is on the protein surface and can freely interact with solvent. To further assess the contributions of the enzyme environment in assisting the catalysis, we have computed the total electrostatic and van der Waals interactions between the QM and MM subsystems for each stationary point along the reaction pathway. As displayed in the Fig. S4, the enzyme and solvent environment, as defined by the MM region in our model, provides strong stabilization to the reaction system, defined here by the QM region. In fact, the stabilization increases from the ES complex through TS-I, INT, TS-II, to EP. This picture is consistent with our perturbation analysis, in which the cationic residues stabilize TS-II more than TS-I as negative charge develops on the phosphate leaving group.

References:

1. Y. Zhu, H. Li, C. Long, L. Hu, H. Xu, L. Liu, S. Chen, D. C. Wang and F. Shao, Mole. Cell **28**, 899 (2007).
2. G. K. Smith, Z. Ke, A. C. Hengge, D. Xu, D. Xie and H. Guo, J. Phys. Chem. B **113**, 15327 (2009).
3. L. Chen, H. Wang, J. Zhang, L. Gu, N. Huang, J.-M. Zhou and J. Chai, Nat. Struc. Mole. Biol. **15**, 101 (2008).
4. D. A. Case, T. Darden, I. Cheatham, T. E., C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, J. Merz, K. M., B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, F. Paesani, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, University of California, San Francisco (2008).
5. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
6. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Proteins **65**, 712 (2006).
7. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
8. N. Homeyer, A. H. C. Horn, H. Lanig and H. Sticht, J. Mole. Model. **12**, 281 (2006).
9. T. Darden, D. York and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
10. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
11. J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, J. Comput. Phys. **23**, 327 (1977).
12. Y. Zhang, T. Lee and W. Yang, J. Chem. Phys. **110**, 46 (1999).
13. Y. Zhang, H. Liu and W. Yang, J. Chem. Phys. **112**, 3483 (2000).
14. Y. Zhang, J. Chem. Phys **122**, 24114 (2005).
15. Y. Zhang, Theor. Chem. Acc. **116**, 43 (2006).
16. G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
17. D. Beeman, J. Comput. Phys. **20**, 130 (1976).
18. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
19. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
20. Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. W. Woodcock III., W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, III., J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, (2006).
21. J. W. Ponder, (2004).

Table SI. Key distances (Å) for stationary points along the reaction path for the eliminylation reaction catalyzed by SpvC

	ES	TS-I	INT	TS-II	EP
$H_{\alpha}(pT)$-N_{ξ}(K136)	2.09±0.09	1.21±0.07	1.06±0.05	1.04±0.04	1.03±0.04
H_{α}(pT)-C_{α}(pT)	1.12±0.05	1.62±0.16	2.39±0.71	2.92±0.35	3.28±0.27
C_{β}(pT)-O_{γ}(pT)	1.45±0.06	1.51±0.07	1.51±0.06	1.84±0.06	2.93±0.12
H_{ε}(H106)-O_{γ}(pT)	2.33±0.40	2.01±0.25	1.93±0.20	1.77±0.16	1.12±0.18
H_{ε}(H106)-N_{ε}(H106)	1.03±0.05	1.03±0.04	1.03±0.04	1.06±0.05	1.53±0.25
C_{α}(pT)-C_{β}(pT)	1.55±0.07	1.52±0.06	1.51±0.06	1.42±0.05	1.34±0.06
C_{α}(pT)-C(pT)	1.53±0.07	1.47±0.06	1.46±0.07	1.44±0.05	1.49±0.06
C_{α}(pT)-N(pT)	1.46±0.06	1.47±0.06	1.46±0.05	1.45±0.05	1.43±0.06
C(pT)-O(pT)	1.24±0.05	1.26±0.05	1.26±0.04	1.27±0.04	1.25±0.05
H(Y158)-O(pT)	1.81±0.16	1.75±0.15	1.79±0.15	1.77±0.15	1.77±0.15
H_2(K104)-O(pT)	1.95±0.19	1.86±0.14	1.83±0.12	1.85±0.12	1.87±0.12

Figure captions:

Figure S1. Two-dimensional minimal energy surface in the proton transfer coordinate, which is defined as the difference between bonds H_{α}-N_{ζ}(Lys136) and H_{α}-C_{α}, and the C_{β}-O_{γ} bond distance. The pink line is the reaction path obtained with RCD method, with the reaction coordinate defined as: $RC = d_{C\beta\text{-}O\gamma} - d_{H\alpha\text{-}N\zeta}$.

Figure S2. Structures of ES, INT, and EP of the wild-type SpvC. The hydrogen bonds of interest are represented in red dash lines.

Figure S3. Individual residue contributions to transition state stabilization. A negative value indicates that the residue favors the reaction and *vice versa*. Note that only the MM residues are included.

Figure S4. Calculated total electrostatic and van der Waals interaction energies between the QM subsystem and its MM environment at ES, TS-I, INT, TS-II and EP, respectively. The averaged energy distribution of ES is shift to zero and the corresponding values are shifted for other stationary points.

Fig. S1

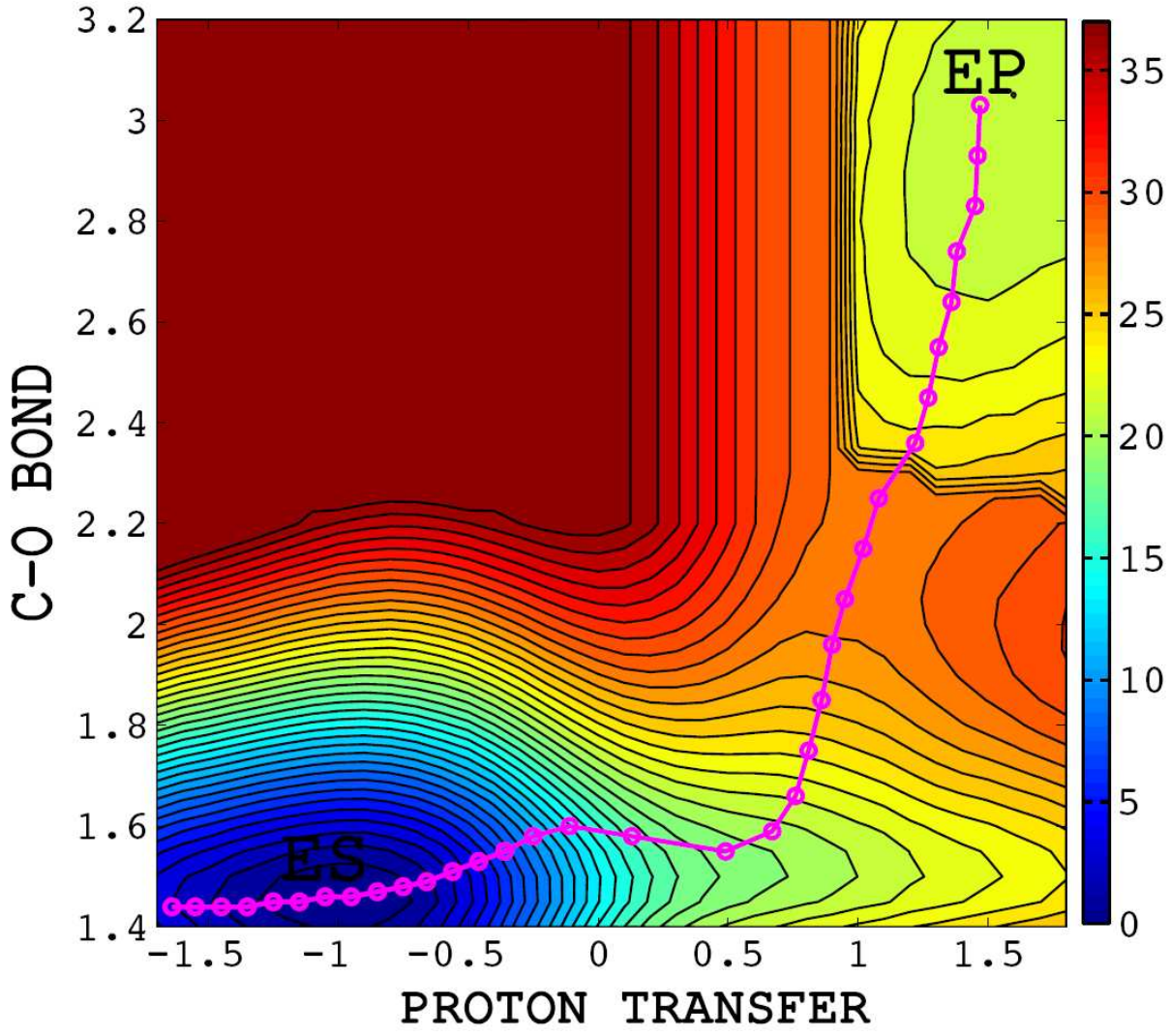

EP
ES
C-O BOND
PROTON TRANSFER
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
-1.5
-1
-0.5
0
0.5
1
1.5
35
30
25
20
15
10
5
0

Fig. S2

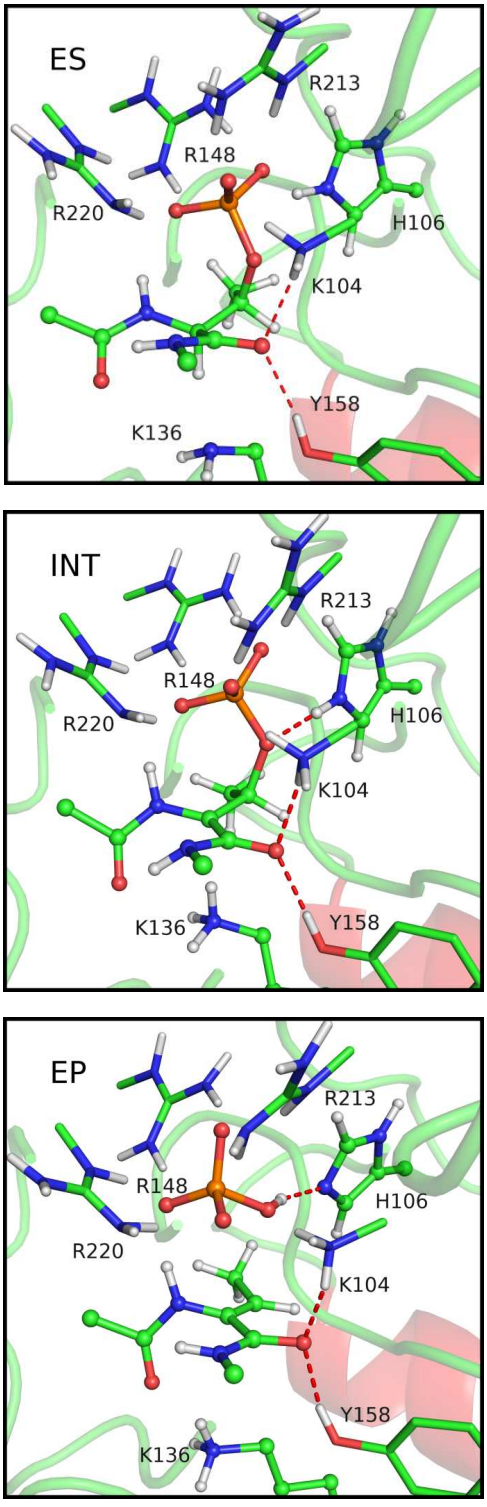

ES
R213
R148
R220
H106
K104
Y158
K136
INT
R213
R148
R220
H106
K104
K136
Y158
EP
R213
R148
H106
R220
K104
Y158
K136

Fig. S3

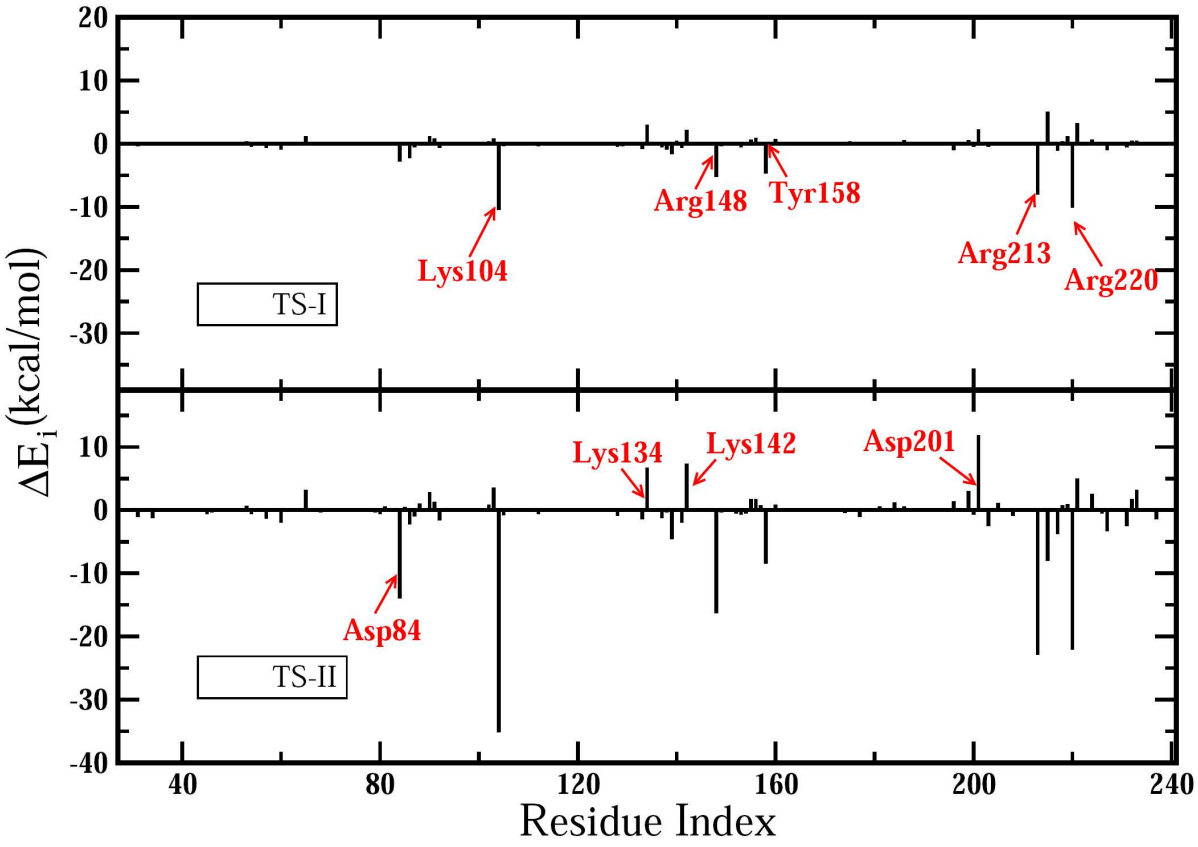
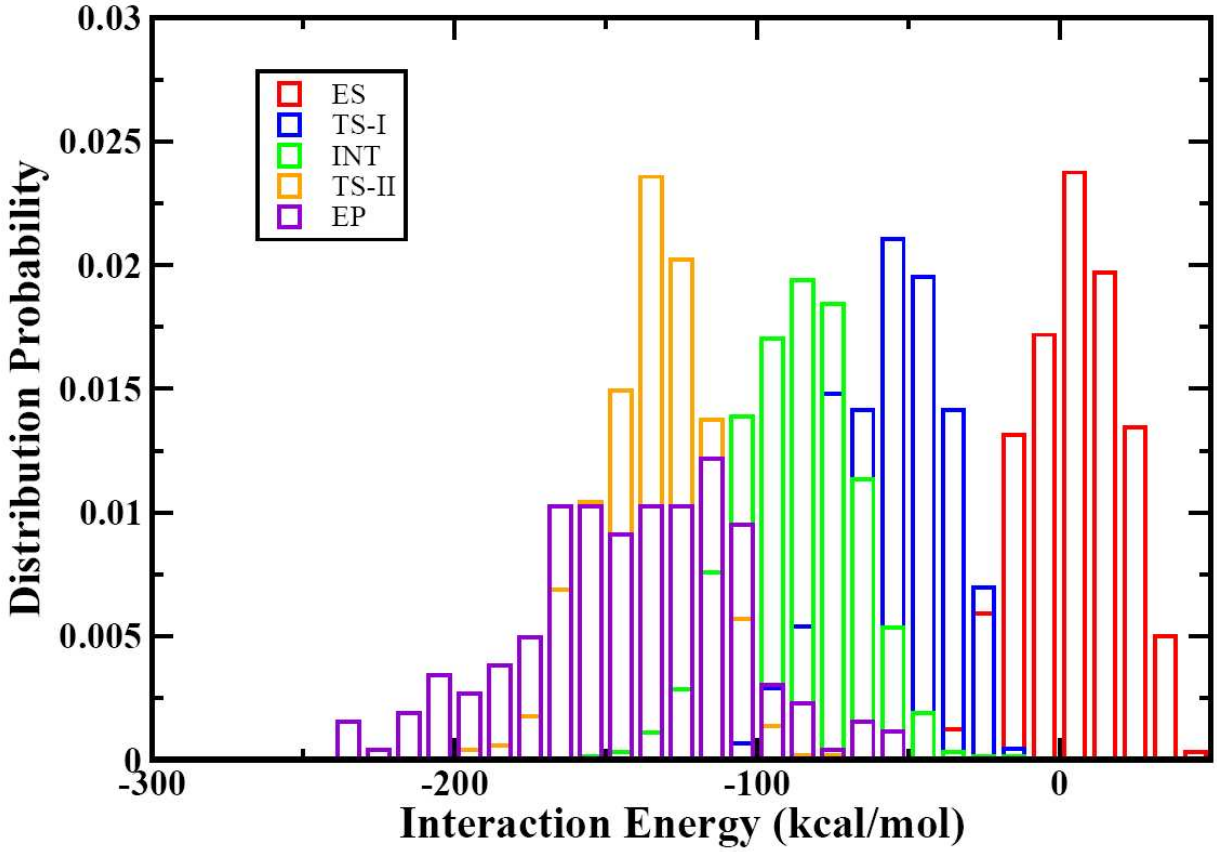

20
10
0
-10
-20
-30
TS-I
Lys104
Arg148
Tyr158
Arg213
Arg220
10
0
-10
-20
-30
-40
TS-II
Asp84
Lys134
Lys142
Asp201
40
80
120
160
200
240
Residue Index
ΔE_i (kcal/mol)

Fig. S4


ES
TS-I
INT
TS-II
EP
Distribution Probability
0.03
0.025
0.02
0.015
0.01
0.005
0
-300
-200
-100
0
Interaction Energy (kcal/mol)