Published in final edited form as:

J Phys Chem B. 2009 September 24; 113(38): 12750–12758. doi:10.1021/jp903173c.

Active Site Cysteine Is Protonated in the PAD4 Michaelis Complex: Evidence from Born-Oppenheimer Ab Initio QM/MM Molecular Dynamics Simulations

Zhihong Ke^{a,b}, Yanzi Zhou^{a,b}, Po Hu^a, Shenglong Wang^a, Daiqian Xie^{b,*}, and Yingkai Zhang^{a,*}

^aDepartment of Chemistry, New York University, New York, NY 10003

^bInstitute of Theoretical and Computational Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210093, China

Abstract

Protein Arginine Deiminase 4 (PAD4) catalyzes the citrullination of the peptidylarginine, and plays a critical role in rheumatoid arthritis (RA) and gene regulation. Understanding its catalytic mechanism is not only of fundamental importance, but also of significant medical interest for the rational design of new inhibitors. By employing on-the-fly Born-Oppenheimer ab initio QM/MM molecular dynamics simulations, we have demonstrated that it is unlikely for the active site cysteine and histidine to exist as a thiolate-imidazolium ion-pair in the PAD4 Michaelis reactant complex. Instead, a substrate-assisted proton transfer mechanism for the deimination reaction step has been characterized: both Cys645 and His471 in the PAD4 active site are neutral prior to the reaction; the deprotonation of Cys645 by the substrate arginine occurs in concert with the nucleophilic addition of the Cys thiolate to C α of the substrate, and leads to a covalent tetrahedral intermediate; then the C α -N η_1 bond cleaves and the resulted ammonia is displaced by a solvent water molecule. The initial deprotonation and nucleophilic attack step is found to be rate-determining. The computed free energy barrier with B3LYP(6-31G*) QM/MM MD simulations and umbrella sampling is 20.9 kcal·mol⁻¹, consistent with the experimental kinetic data. During the deimination, His471 plays an important role in stabilizing transition state through the formation of the hydrogen bond with the guanidinium group. Our current studies further demonstrated the viability and strength of the ab initio QM/MM molecular dynamics approach in simulating enzyme reactions.

Keywords

enzyme reaction; mechanism; deimination; protein arginine deiminase 4; free energy barrier

1 Introduction

Protein arginine deiminase 4 (PAD4) is a calcium-dependent enzyme that catalyzes the hydrolysis of peptidylarginine residue to form citrulline residue and ammonia in a post-translational modification process [1,2]. It plays a pivotal role in human gene regulation by deiminating multiple transcriptional regulators, such as p300, and histones H2A, H3, H4 [1, 3-6], and the dysregulation of PAD4 has been linked to the onset and progression of the autoimmune disorder relevant to rheumatoid arthritis [7-10]. To elucidate the catalytic

* Address corresponding to these authors. dqxie@nju.edu.cn, yingkai.zhang@nyu.edu.

mechanism of PAD4 is not only of fundamental importance, but also of significant medical interest for the rational design of new mechanism-based inhibitors.

PAD4 is a member of guanidino-group modifying enzyme superfamily (GMSF) [11], which includes arginine deiminase (ADI) [12-19] and dimethylarginine dimethylaminohydrolase (DDAH) [20-22]. In spite of considerable sequence diversity among GMSF enzymes, their overall structural fold of the catalytic domain and active site architecture are quite similar. The basic catalytic motif, a Cys-His-Asp(Glu) triad, is conserved among all GMSF members, and is located at the center of the catalytic domain, a structure of five $\beta\beta\alpha\beta$ modules arranged circularly in a pseudo-five-fold symmetric structure. For PAD4, besides the catalytic domain at its C-terminal, it also consists of an N-terminal domain which is far away from the active site and does not directly affect its catalytic activity [23,24] (see Fig. 1). The active site of PAD4 is enriched in charged and polar residues, including: Asp350, His471, Asp473, His644 and Cys645. Structural analysis and site-directed mutagenesis studies have indicated that there are totally five calcium ions binding in activated PAD4, among which the two Ca^{2+} coordinating to the residues of the C-terminal domain are crucial for catalysis whereas the other three have almost no effect on the enzyme activity [23,24].

Based on extensive structural and biochemical studies [14,15,20,22,25], it has been generally suggested that PAD4 adopts a similar two-stage catalytic mechanism like other GMSF members [11,14,15,22,26,27]: in the initial deimination stage, the nucleophilic attack of the active site Cys645 to the guanidinium carbon of the arginine residue is followed by the cleavage of the $\text{C}_\zeta\text{-N}_{\eta 1}$ bond to release ammonia; then in the second reaction stage, the formed covalent thiouronium intermediate is hydrolyzed to generate the product citrulline.

Although the proposed reaction scheme is reasonable in general, many detailed mechanistic questions remain elusive. Among them, the most controversial and puzzling issue about the enzyme mechanism for PAD4, as well as for other GMSF members, is the protonation state of the active site cysteine in the Michaelis reactant complex. Since the thiol group itself is a poor nucleophile group and no base group is located close enough to directly deprotonate the active site Cys in all structures solved for GMSF enzymes, it has been widely believed that the active cysteine (Cys645 in PAD4) is deprotonated while the active histidine (His471 in PAD4) is protonated in the enzyme-substrate complex (state I as shown in Scheme 1 of Fig. 2) [28, 29]. In the first step of the reaction, the thiolate group directly attacks while the protonated histidine serves as the general acid, as shown in Fig. 2. However, at the apo-enzyme state, it has been found that two active site residues have pK_a values of 8.2 and 7.3 [27], while the active Cys in ADI has a pK_a of 9.6 [19]. Considering that the intrinsic pK_a values of cysteine and histidine residues are 8.3 and 6.0 respectively, and PAD4 as well as other members of GMSF are activated at a neutral or acidic pH value, it brings the fundamental question regarding how Cys645 can be deprotonated in PAD4. Thompson and his co-workers suggested a reverse protonation mechanism [27] for PAD4 that the protonated His471 and the deprotonated Cys645 directly form an ion-pair in the free and active forms of PAD4 [27], analogous to cysteine proteases [28,29]. However, the distance between C_β atom of the cysteine and N_δ atom of the histidine is at least 6 Å in PAD4, and the corresponding distance in ADI and DDAH, two other GMSF enzymes, are even larger. For DDAH [22] and ADI [19], a substrate-assisted cysteine deprotonation mechanism has been suggested, in which the active Cys is protonated in the apo-enzyme, and the proton is lost due to the binding of the positive substrate.

As shown in Fig. 2, the other mechanism hypothesis (Scheme 2) is that Cys645 and His471 are both neutral in the PAD4 Michaelis complex, which is referred as the state N. To overcome the apparent difficulty that no base group has been located to be close enough to directly deprotonate the active site Cys, an substrate-assisted proton transfer mechanism has been proposed [11,13,30] in which the thiol proton is transferred to imidazole of histidine via a

substrate guanidino nitrogen. An alternative hypothesis has also been mentioned that the nearby aspartic acid residue can be the proton-accepting group instead of the histidine [13].

Since experimental means alone are very difficult to unambiguously distinguish among these different mechanistic hypotheses, there is of significant interest in simulating GMSF enzymes. However, ab initio quantum mechanical studies on a model complex of the enzyme active site [30,31] neglect the heterogeneous enzyme environment, and its applicability in modeling enzyme reactions has been well recognized to be very limited. Very recently, Guo and his co-workers have employed ab initio QM/MM methods to map out reaction potential energy profiles for the deimination step of ADI, and have explored both schemes in Fig. 2. They found that the reaction energy barriers for scheme 1 and scheme 2 are 2 and 33 kcal·mol⁻¹, respectively [19]. Thus, it was concluded that the calculations supported the Scheme 1, in which both the active site cysteine and histidine are charged in the reactant complex. However, 2 kcal·mol⁻¹ barrier would be typically too small for an enzyme reaction. Meanwhile, it should be noted that both reaction schemes lead to the same intermediate, so that their calculations would support that the neutral state of the ADI would be more stable than the ionic state.

In order to shed more light on these important mechanistic issues, we have carried out Born-Oppenheimer ab initio QM/MM molecular dynamics simulations on both probable reactant states of PAD4, i.e., the state I and N as shown in Fig. 2, and determined the free energy reaction profile for the initial deimination reaction step. Our employed approach represents the state-of-the-art method to simulate enzyme reactions: it provides an ab initio description of chemical bond formation and breaking process, properly and explicitly models the heterogeneous enzyme environment including both protein and solvent water molecules, and takes accounts of dynamics of enzyme reaction center and its environment on an equal footing. Our simulations have indicated that the state I, in which Cys645 and His471 are charged, is not stable; instead, both of them are neutral in the PAD4 active site prior to the reaction, and PAD4 employs an alternative substrate-assisted nucleophilic attack mechanism for the deimination reaction step.

2 Methods

To simulate enzyme reactions, extensive sampling on a reasonably accurate potential energy surface is needed to obtain reliable results. The free energy changes associated with the reactions are not only better-defined for such systems, but also characterize the reactions better than potential energies. Here our theoretical approaches center on the Born-Oppenheimer molecular dynamics simulations with ab initio QM/MM potential [32-43] and the umbrella sampling method [44-46]: at each time step, the atomic forces as well as the total energy of the QM/MM system are calculated with the B3LYP(6-31G*) QM/MM approach on-the-fly, and Newton equations of motion are integrated. From a series of biased simulations, the potential of mean force along the reaction coordinate is obtained with the weighted histogram analysis method (WHAM). This direct ab initio QM/MM MD approach and the corresponding simulation protocol has been successfully applied to elucidate the catalytic power of SET7/9 [47], provide insights into the methylation state specificity of histone lysine methylation [48], and characterize the Sir2 catalyzed nicotinamide cleavage reaction [49].

The initial structure of the PAD4 enzyme-substrate complex was prepared based on the crystal structure 2DEY, which was an inactive C645A mutant in complex with histone H4 N-terminal tail including Arg3 in 2.25 Å resolution [24]. Both ionic and neutral models of PAD4, which differ in the protonation states for Cys645 and His471 as shown in Fig. 2, were prepared with the same procedure and equilibrated with the same simulation protocol. Since the N-terminal domain of PAD4 is far away from the active site and it does not affect the PAD4 catalytic function [23,24] (see Fig. 1), residues Met1 to Ala294 as well as their surrounding water and

ions were removed to reduce computational cost. Ala645 was mutated back to the Cys residue, and hydrogen atoms were added by Leap in the Amber package [50]. For the PAD4 at the neutral state, Cys645 is protonated and His471 is neutral as HIE, while for that at the charged state, Cys645 is deprotonated and His471 is charged. The local hydrogen bond network has been carefully checked in determining the protonation states of all other histidine residues: His644 has been identified to be protonated as HIP since it would form two hydrogen bonds with Asp473 and a crystal water, while all the rest histidine residues are neutral. The resulting complex was neutralized by adding 7 Cl^- counter-ions and solvated into a periodic TIP3P rectangular water box of $\sim 69 \times 83 \times 71 \text{ \AA}^3$, with a buffer distance of 10 $\text{\AA}$ between each wall and the closest atom in each direction. This led to a simulation system of ~ 40600 atoms for either PAD4 simulation system.

Minimizations and MD simulations with periodic boundary condition were conducted with AMBER molecular dynamics package, employing AMBER99SB force field [51,52] and TIP3P model [53] for water molecules. An 8 $\text{\AA}$ cutoff was introduced for non-bonding interactions, updating the neighbor pair list every 10 steps. Particle Mesh Ewald (PME) method [54,55] was employed to treat electrostatic interactions, and the SHAKE algorithm [56] was used to constrain all bonds involving hydrogens. A time step of 1 fs was used. The added solvent molecules and ions were first minimized and then relaxed by a 60 ps NVT, and a 40 ps NPT MD simulations with the heavy atoms of protein and retained crystal water being restrained by a force constant of $50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$. The resulting system was energy refined by releasing the restraint on protein gradually, and finally a 2 ns un-restrained MD simulation was carried out. The trajectories for both models are quite stable. The coordination shell of two Ca^{2+} ions has also kept quite well during the simulations. With a snapshot after 1 ns, the QM/MM model was obtained by removing the water outside of 27 $\text{\AA}$ of C_ζ of the substrate. The resulting ionic and neutral systems had 9721 and 9677 atoms, respectively.

The QM/MM partition is illustrated in Figure 3. In all previously proposed mechanisms, possible fragments directly involved in the reaction include side-chains of Cys645, His471, Asp473 and the substrate Arg, which have been described by B3LYP functional with 6-31G(d) basis set. The QM/MM boundary was treated with the pseudobond approach [57,58]. All other atoms are described by AMBER99SB force field [51,52] and TIP3P model [53] for water molecules. Each prepared QM/MM system was first minimized by QM/MM calculations and then MM subsystem was further equilibrated by carrying out a 500 ps MD simulation with MM force field. The resulting snapshot was used as the starting structure for the ab initio QM/MM MD simulations without any restraints for atoms within 20 $\text{\AA}$ of C_ζ of the substrate.

Ab initio QM/MM molecular dynamics simulations with umbrella sampling have been employed to determine the free energy profile for the enzyme reaction catalyzed by PAD4. Starting from the QM/MM minimized structure, a rough minimum energy path was first determined with an iterative restrained minimization procedure and reaction coordinate driving (RCD) method [59,60]. For each determined structure along the path, a 500 ps MD simulation with MM force field was carried out to equilibrate the MM subsystem. The resulting snapshot was used as the starting structure for ab initio QM/MM MD simulation with umbrella sampling [44-46]. For each window, the total potential energy was biased with a harmonic potential of $40 \sim 100 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ centered on corresponding value of the reaction coordinate. The configurations were collected for 20 ps for data analysis after an equilibration period of 10 ps. The probability distributions along the defined reaction coordinate were determined for each window and pieced together with the weighted histogram analysis method (WHAM) [61-63] to calculate the free energy profile. In average, it took about 12 hours in wall time to run 1 ps B3LYP/6-31G(d) QM/MM MD simulation on a quad-core 2.33 GHz Intel Xeon cluster using 16 cores for this system.

All our QM/MM simulations have been carried out with modified versions of Gaussian 03 [64], QCHEM [65] and TINKER [66] programs. The spherical boundary condition has been applied so that the atoms outside of 20 Å of C_γ of the substrate were fixed. A cutoff of 12 Å for van der Waals interactions, and a cutoff of 18 Å were employed for electrostatic interactions among MM atoms. There is no cutoff for electrostatic interactions between QM and MM atoms. Berendsen thermostat method [67] has been used to control the system temperature at 300 K.

3 Results and Discussions

3.1 Michaelis Reactant Complex

The most controversial issue about PAD4 is the protonation states of active site residues in the Michaelis reactant complex. Two probable reactant states have been proposed: one is the state I in which Cys645 is deprotonated and His471 is charged (Scheme 1, Fig. 2), and the other is the state N in which both Cys645 and His471 are neutral (Scheme 2, Fig. 2). Correspondingly, we have prepared two models based on the crystal structure 2DEY with the same procedure, and they have been investigated with the same simulation protocol. Although both models are stable in the 2 ns molecular dynamics simulations with the molecular mechanical force field, only the active site of the neutral model is stable in the 35 ps B3LYP(6-31G*) QM/MM MD simulation, as illustrated in Fig. 4. For the ionic model, its active site undergoes a significant change during the QM/MM MD simulation, which leads to the disruption of the enzyme-substrate complex.

As shown in Fig. 4, in the crystal structure, the substrate is tightly held by the four hydrogen bonds with Asp350 and Asp473, and Asp473 also forms a hydrogen bond with His644. This overall hydrogen bond network in the active site is well conserved throughout the 35 ps ab initio QM/MM MD simulation of the neutral model, which is illustrated as State N in Fig. 4. The sulfur atom of Cys645 is positioned right under C_γ of the substrate with an average distance of 3.72 ± 0.15 Å. Overall, for the neutral model, we can see that the active site residues are orientated well for the proton transfer from the thiol to the substrate and the subsequent nucleophilic addition.

In contrast, for the ionic model, the hydrogen network in the active site falls apart during the 35 ps ab initio QM/MM MD simulation, which is illustrated as State I in Fig. 4. The substrate has drifted away from the active site during the first few pico-seconds and in consequence, the hydrogen bonds between the substrate and two aspartic acid, Asp350 and Asp473, are broken. Meanwhile, two water molecules have entered into the active site, forming hydrogen bonds with the S⁻ atom of Cys645. In addition, we have also performed QM/MM MD simulations with two entering water molecules included in the QM sub-system, the leaving of the substrate and the entering of water molecules still occur. All these indicate that the ionic model of the PAD4 complex is not stable in the ab initio QM/MM MD simulation.

In order to further compare the stability of these two states, Born-Oppenheimer ab initio QM/MM MD simulations with umbrella sampling method have been employed to calculate the free energy profile corresponding to the transformation from the neutral state to the ionic state. As shown in Fig. 5, starting from the neutral state, the free energy curve is first smooth and uphill until the reaction coordinate of 0.60 Å, with the height of $13.3 \text{ kcal}\cdot\text{mol}^{-1}$, and then it decreases until reaching the ionic state which has a free energy $6.0 \text{ kcal}\cdot\text{mol}^{-1}$ higher than the neutral form. For the snapshots corresponding to the highest point along the free energy profile, HS is partially bonded with N_{H1} of the substrate with a distance of 1.19 ± 0.08 Å, and the distance between HS and S of Cys645 becomes 1.77 ± 0.09 Å, as illustrated in Fig. 6. For subsequent umbrella windows after the RC of 0.60 Å, it is found that substrate starts to drift away and water molecules move into the active site. A snapshot corresponding to the RC of 1.5 Å is also illustrated in Fig. 6, which has similar features as the State I in Fig. 4. The results are quite

consistent with ab initio QM/MM MD simulations of the ionic model of the PAD4 presented above. It should be noted that due to the entering of MM water molecules into the active site, the free energy profile after the RC of 0.6 Å is not quantitatively reliable, thus it has been represented with dashes. Nevertheless, these results further confirm that it is unlikely for the active site Cys645 and His471 to be charged in the PAD4 Michaelis reactant complex.

3.2 Deimination Reaction Mechanism

Our studies so far have indicated that only the neutral state of PAD4 reactant complex is stable, in which both Cys645 and His471 are neutral. The next essential task is to computationally characterize the deimination reaction catalyzed by PAD4. By employing ab initio QM/MM MD simulations with umbrella sampling method, we have determined the free energy profile for the deimination reaction catalyzed by PAD4, as shown in Fig. 7. The calculated free energy barrier for the deimination step is 20.9 kcal·mol⁻¹, which is in agreement with the experimental value of 17.0 kcal·mol⁻¹ estimated from the k_{cat} value of 6.55 s⁻¹ for PAD4 [26]. Our simulations suggest an alternative substrate-assisted proton transfer mechanism [11,13,30] for the PAD4 catalyzed deimination reaction. It consists of two reaction steps: in the first step, the nucleophilic addition of the Cys thiol to the C $_{\alpha}$ of the substrate occurs in concert with the deprotonation of Cys645 by the substrate arginine, which leads to a covalent tetrahedral intermediate; then in the second step, the C $_{\alpha}$ -N $_{\eta 1}$ bond is cleaved to form the thiouronium intermediate and an ammonia molecule. The initial deprotonation and nucleophilic attack step is rate-determining, as shown in Fig. 7.

The characterized reactant, transition states and intermediates are illustrated in Fig. 8 and the key geometric parameters are listed in table 1. In the reactant complex, the guanidinium group of the substrate forms four hydrogen bonds with Asp473 and Asp350, but it is not directly hydrogen bonded with His471. From the reactant to the transition state I, the proton of thiol group transfers to the N $_{\eta 1}$ atom and the ionized S atom approaches C $_{\alpha}$ atom of substrate. At the transition state I, the guanidinium moiety is not planar, which indicates the conjugate system of the guanidinium is broken. For the hydrogen bond network, one hydrogen bond between Asp350 and the substrate becomes broken while the other three are conserved, in which the hydrogen bond distance between N $_{\eta 1}$ of the substrate and O $_{\delta 2}$ of Asp473 has been shortened from 2.82 ± 0.11 Å in the reactant to 2.71 ± 0.12 Å at TS I. Meanwhile, a new hydrogen bond has been formed between N $_{\eta 1}$ of the substrate and N $_{\delta}$ of His471, whose distance has changed from 3.34 ± 0.16 Å in the reactant to 2.80 ± 0.11 Å at the TS I. In addition, during the ab initio QM/MM MD simulation of the TS I, it is found that the protons of the substrate guanidinium group can partially and transiently transfer to either His471 or Asp473. Thus two representative configurations have been illustrated in Fig.8. To better characterize such transient proton transfer, we have monitored both H $_{11}$ -N $_{\eta 1}$ and H $_{12}$ -N $_{\eta 1}$ bond lengths during a 50 ps QM/MM MD simulation for the window of the transition state, as illustrated in Fig. 9. H $_{11}$ and H $_{12}$ refer to two protons hydrogen bonded with His471 and Asp473 respectively. The significant elongation of H $_{11}$ -N $_{\eta 1}$ or H $_{12}$ -N $_{\eta 1}$ indicates the proton transfer from the substrate to His471 or Asp473, respectively. Fig. 9 clearly demonstrates the partial and transient proton transfer between the substrate guanidinium group and His471 or Asp473 at the transition state. These results suggest that both His471 and Asp473 play essential roles to stabilize the transition state through the formation of hydrogen bonds with the transition state. which is very consistent with the site-directed mutagenesis results regarding His471 and Asp473 in PAD4 [27]. For Asp350, our results suggest that its main catalytic role is likely to facilitate the substrate binding rather than to stabilize the transition state. With the further shortening of the S-C $_{\alpha}$ bond, the tetrahedral intermediate (Int. I) is formed, which is 12.3 kcal·mol⁻¹ higher in free energy than the reactant. As shown in Fig. 8, the C $_{\alpha}$ becomes sp³ hybridization at the tetrahedral intermediate, with the S-C $_{\alpha}$ being 1.93 ± 0.06 Å while C $_{\alpha}$ -N $_{\epsilon}$, C $_{\alpha}$ -N $_{\eta 1}$ and C $_{\alpha}$ -N $_{\eta 2}$ being elongated to 1.40 ± 0.03, 1.58 ± 0.06 and 1.42 ± 0.03 Å, respectively. For the intermediate I,

its hydrogen bond network at the active site is similar to the TS I but no transient proton transfer between the guanidinium group and His471 or Asp473 has been observed throughout the ab initio QM/MM MD simulations.

Starting from the Int. I, the $C_{\alpha}-N_{\eta 1}$ bond breaks to yield the thiouronium intermediate and an ammonia molecule. The activation barrier for this step is $6.0 \text{ kcal}\cdot\text{mol}^{-1}$, which is much lower than the first step and is not rate-limiting. From the tetrahedral intermediate (Int. I) to TS II, the $C_{\alpha}-N_{\eta 1}$ bond length has changed from 1.58 ± 0.06 to $2.10 \pm 0.01 \text{ \AA}$ but the overall hydrogen bond network remains similar. When the $C_{\alpha}-N_{\eta 1}$ further gets longer to yield the thiouronium intermediate, the ammonia molecule is formed. At the thiouronium intermediate, C_{α} returns to be sp^2 hybridization, and as a consequence, the $C_{\alpha}-N_{\epsilon}$, $C_{\alpha}-N_{\eta 2}$ and $C_{\alpha}-S$ bonds are shortened to 1.34 ± 0.02 , 1.31 ± 0.02 and $1.79 \pm 0.03 \text{ \AA}$, respectively. The thiouronium moiety form three hydrogen bonds with Asp350, His471 and Asp473, respectively.

For the thiouronium intermediate to be further hydrolyzed, the ammonium molecule needs to diffuse into the bulk solvent first. We have carried out MM molecular dynamics simulations with umbrella sampling method to explore this ammonia release process. Our simulations indicate that the leaving of ammonia from the active site encounters a low barrier, which is about $2.7 \text{ kcal}\cdot\text{mol}^{-1}$, given in Fig. 10. The energy fluctuation is probably due to the forming and breaking of the weak hydrogen bond between NH_3 and the nearby water molecule. PAD4 undergoes small change during the diffusion, and the largest RMSD for the backbone is below $1.5 \text{ \AA}$. The channel employed for the ammonium diffusion is found to be distinct from the one used for the substrate binding, as shown in Fig. 11. Consequently, the leaving of the ammonia molecule would not affect the binding of the thiouronium intermediate in the active site. By analyzing simulation trajectories, it is found that after NH_3 is diffused, a water molecule comes into the active site and forms hydrogen bonds with His471 and Asp473. As shown in Fig. 11, the entering water molecule is orientated well to make an attack to the C_{α} atom for the subsequent hydrolysis reaction. This sets the stage for further theoretical characterization of second hydrolysis stage of the PAD4 catalyzed reaction.

4 Conclusion

With Born-Oppenheimer ab initio QM/MM MD simulation, which is the state-of-the-art approach to simulate enzyme reactions, we have characterized the Michaelis reactant complex and determined the free energy profile of the initial deimination reaction for PAD4. Our simulations provided strong evidence that the active site Cys645 is protonated in the PAD4 reactant complex with His471 being singly protonated, and PAD4 employs a substrate-assisted proton transfer mechanism for the deimination reaction step. During the ab initio QM/MM MD simulations of the characterized transition state, the partial and transient transfer of the protons from the guanidinium group to either His471 or Asp473 has been observed, which indicates important roles of His471 and Asp473 in transition state stabilization. To provide such detailed mechanistic insights into inner workings of PAD4 is not only of fundamental importance, but also of significant medical interest for the rational design of new mechanism-based inhibitors. Our current investigations also set the stage for investigating the second hydrolysis step of the PAD4 catalyzed reaction as well as for revisiting another important GMSF enzyme ADI.

Acknowledgments

This work was partly supported by NIH (R01-GM079223), NSF (CHE-CAREER-0448156) and the China Scholarship Council. D. X. was supported by the National Natural Science Foundation of China (Grant Nos. 20725312 and 20533060) and the Ministry of Science and Technology (2007CB815201). We thank NYU-ITS for providing computational resources.

References

- [1]. Hagiwara T, Nakashima K, Hirano H, Senshu T, Yamada M. *Biochem. and Biophys. Res. Commun* 2002;290:979–983. [PubMed: 11798170]
- [2]. Vossenaar ER, Zendman AJW, van Venrooij WJ, Pruijn GJM. *BioEssays* 2003;25:1106–1118. [PubMed: 14579251]
- [3]. Nakashima K, Hagiwara T, Yamada M. *J. Biol. Chem* 2002;277:49562–49568. [PubMed: 12393868]
- [4]. Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, Sonbuchner LS, McDonald CH, Cook RG, Dou Y, Roeder RG, Clarke S, Stallcup MR, Allis CD, Coonrod SA. *Science* 2004;306:279–283. [PubMed: 15345777]
- [5]. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, Schneider R, Gregory PD, Tempst P, Bannister AJ, Kouzarides T. *Cell* 2004;118:545–553. [PubMed: 15339660]
- [6]. Hagiwara T, Hidaka Y, Yamada M. *Biochemistry* 2005;44:5827–5834. [PubMed: 15823041]
- [7]. Suzuki A, Yamada R, Chang XT, Tokuhira S, Sawada T, Suzuki M, Nagasaki M, Nakayama-Hamada M, Kawaida R, Ono M, Ohtsuki M, Furukawa H, Yoshino S, Yukioka M, Tohma S, Matsubara T, Wakitani S, Teshima R, Nishioka Y, Sekine A, Iida A, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. *Nat. Genet* 2003;34:395–402. [PubMed: 12833157]
- [8]. Vossenaar ER, van Venrooij WJ. *Arthritis Res. Ther* 2004;6:107–111. [PubMed: 15142259]
- [9]. Vossenaar ER, Despres N, Lapointe E, van der Heijden A, Lora M, Senshu T, van Venrooij WJ, Menard HA. *Arthritis Res. Ther* 2004;6:R142–R150. [PubMed: 15059278]
- [10]. Foulquier C, Sebbag M, Clavel C, Chapuy-Regaud S, Al Badine R, Mechin MC, Vincent C, Nachat R, Yamada M, Takahara H, Simon M, Guerrin M, Serre G. *Arthritis Rheum* 2007;56:3541–3553. [PubMed: 17968929]
- [11]. Shirai H, Blundell TL, Mizuguchi K. *Trends Biochem.Sci* 2001;26:465–468. [PubMed: 11504612]
- [12]. Lu XF, Galkin A, Herzberg O, Dunaway-Mariano D. *J. Am. Chem. Soc* 2004;126:5374–5375. [PubMed: 15113205]
- [13]. Galkin A, Kulakova L, Sarikaya E, Lim K, Howard A, Herzberg O. *J. Biol. Chem* 2004;279:14001–14008. [PubMed: 14701825]
- [14]. Das K, Butler GH, Kwiatkowski V, Clark AD, Yadav P, Arnold E. *Structure* 2004;12:657–667. [PubMed: 15062088]
- [15]. Galkin A, Lu XF, Dunaway-Mariano D, Herzberg O. *J. Biol. Chem* 2005;280:34080–34087. [PubMed: 16091358]
- [16]. Lu XF, Li L, Feng XH, Wu Y, Dunaway-Mariano D, Engen JR, Mariano PS. *J. Am. Chem. Soc* 2005;127:16412–16413. [PubMed: 16305225]
- [17]. Lu XF, Li L, Wu R, Feng XH, Li ZM, Yang HY, Wang CH, Guo H, Galkin A, Herzberg O, Mariano PS, Martin BM, Dunaway-Mariano D. *Biochemistry* 2006;45:1162–1172. [PubMed: 16430212]
- [18]. Li L, Li ZM, Chen DQ, Lu XF, Feng XH, Wright EC, Solberg NO, Dunaway-Mariano D, Mariano PS, Galkin A, Kulakova L, Herzberg O, Green-Church KB, Zhang L. *J. Am. Chem. Soc* 2008;130:1918–1931. [PubMed: 18205354]
- [19]. Li L, Li ZM, Wang CH, Xu DG, Mariano PS, Guo H, Dunaway-Mariano D. *Biochemistry* 2008;47:4721–4732. [PubMed: 18366187]
- [20]. Stone EA, Person MD, Costello NJ, Fast W. *Biochemistry* 2005;44:7069–7078. [PubMed: 15865451]
- [21]. Stone EM, Schaller TH, Bianchi H, Person MD, Fast W. *Biochemistry* 2005;44:13744–13752. [PubMed: 16229464]
- [22]. Stone EM, Costello AL, Tierney DL, Fast W. *Biochemistry* 2006;45:5618–5630. [PubMed: 16634643]
- [23]. Arita K, Hashimoto H, Shimizu T, Nakashima K, Yamada M, Sato M. *Nat. Struct. Mol. Biol* 2004;11:777–783. [PubMed: 15247907]
- [24]. Arita K, Shimizu T, Hashimoto H, Hidaka Y, Yamada M, Sato M. *Proc. Natl. Acad. Sci. U. S. A* 2006;103:5291–5296. [PubMed: 16567635]

- [25]. Luo Y, Arita K, Bhatia M, Knuckley B, Lee YH, Stallcup MR, Sato M, Thompson PR. *Biochemistry* 2006;45:11727–11736. [PubMed: 17002273]
- [26]. Kearney PL, Bhatia M, Jones NG, Yuan L, Glascock MC, Catchings KL, Yamada M, Thompson PR. *Biochemistry* 2005;44:10570–10582. [PubMed: 16060666]
- [27]. Knuckley B, Bhatia M, Thompson PR. *Biochemistry* 2007;46:6578–6587. [PubMed: 17497940]
- [28]. Ma S, Devi-Kesavan LS, Gao J. J. *Am. Chem. Soc* 2007;129:13633–13645. [PubMed: 17935329]
- [29]. Mladenovic M, Fink RF, Thiel W, Schirmeister T, Engels B. J. *Am. Chem. Soc* 2008;130:8696–8705. [PubMed: 18557615]
- [30]. Wang CH, Xu DG, Zhang LD, Xie DQ, Guo H. J. *Phys. Chem. B* 2007;111:3267–3273. [PubMed: 17388453]
- [31]. M. Leopoldini TM, Toscano M. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling* 2008;120:459–466.
- [32]. Warshel A, Levitt M. J. *Mol. Biol* 1976;103:227–249. [PubMed: 985660]
- [33]. Singh UC, Kollman PA. *J. Comput. Chem* 1986;7:718–730.
- [34]. Field MJ, Bash PA, Karplus M. J. *Comput. Chem* 1990;11:700–733.
- [35]. Gao J, Truhlar DG. *Annu. Rev. Phys. Chem* 2002;53:467–505. [PubMed: 11972016]
- [36]. Cui Q, Karplus M. *Adv. Protein Chem* 2003;66:315–372. [PubMed: 14631822]
- [37]. Shurki A, Warshel A. *Adv. Protein Chem* 2003;66:249–313. [PubMed: 14631821]
- [38]. Friesner RA, Guallar V. *Annu. Rev. Phys. Chem* 2005;56:389–427. [PubMed: 15796706]
- [39]. Mulholland AJ. *Drug Discov. Today* 2005;10:1393–1402. [PubMed: 16253878]
- [40]. Zhang Y. *Theor. Chem. Acc* 2006;116:43–50.
- [41]. Senn HM, Thiel W. *Top. Curr. Chem* 2007;268:173–290.
- [42]. Hu H, Yang W. *Annu. Rev. Phys. Chem* 2008;59:573–601. [PubMed: 18393679]
- [43]. Senn HM, Thiel W. *Angewandte Chemie International Edition* 2009;48:1198–1229.
- [44]. Patey GN, Valleau JP. *J. Chem. Phys* 1975;63:2334–2339.
- [45]. Boczek EM, Brooks CL. *J. Phys. Chem* 1993;97:4509–4513.
- [46]. Roux B. *Comput. Phys. Commun* 1995;91:275–282.
- [47]. Wang S, Hu P, Zhang Y. J. *Phys. Chem. B* 2007;111:3758–3764. [PubMed: 17388541]
- [48]. Hu P, Wang S, Zhang Y. J. *Am. Chem. Soc* 2008;130:3806–3813. [PubMed: 18311969]
- [49]. Hu P, Wang S, Zhang Y. J. *Am. Chem. Soc* 2008;130:16721–16728. [PubMed: 19049465]
- [50]. Case, DA.; Darden, TA.; T. E. Cheatham, I.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Merz, KM.; Pearlman, DA.; Crowley, M.; Walker, RC.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, KF.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, DH.; Schafmeister, C.; Ross, WS.; Kollman, PA. *AMBER 9*. University of California; San Francisco: 2006.
- [51]. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J. Am. Chem. Soc* 1995;117:5179–5197.
- [52]. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. *Proteins* 2006;65:712–725. [PubMed: 16981200]
- [53]. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J. Chem. Phys* 1983;79:926–935.
- [54]. Darden T, York D, Pedersen L. *J. Chem. Phys* 1993;98:10089–10092.
- [55]. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. *J. Chem. Phys* 1995;103:8577–8593.
- [56]. Ryckaert JP, Ciccotti G, Berendsen HJC. *J. Comput. Phys* 1977;23:327–341.
- [57]. Zhang Y, Lee T, Yang W. *J. Chem. Phys* 1999;110:46–54.
- [58]. Zhang Y. *J. Chem. Phys* 2005;122:024114. [PubMed: 15638579]
- [59]. Zhang Y, Liu H, Yang W. *J. Chem. Phys* 2000;112:3483–3492.
- [60]. Williams IH, Maggiora GM. *Theochem. J. Mol. Struct* 1982;6:365–378.
- [61]. Alan M, Ferrenberg DPL, Swendsen RH. *Phys.Rev.E* 1988;51:5092–5100.

- [62]. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. *J. Comput. Chem* 1992;13:1011–1021.
- [63]. Souaille M, Roux B. *Comput. Phys. Commun* 2001;135:40–57.
- [64]. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr.; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03, Revision D.01. Gaussian, Inc.; Wallingford, CT: 2004.
- [65]. Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, ST.; Gilbert, ATB.; Slipchenko, LV.; Levchenko, SV.; O'Neill, DP.; Lochan, RC.; Wang, T.; Beran, GJO.; Besley, NA.; Herbert, JM.; Lin, CY.; Voorhis, TV.; Chien, SH.; Sodt, A.; Steele, RP.; Rassolov, VA.; Maslen, PE.; Korambath, PP.; Adamson, RD.; Austin, B.; Baker, J.; Byrd, EFC.; Dachsel, H.; Doerksen, RJ.; Dreuw, A.; Dunietz, BD.; Dutoi, AD.; Furlani, TR.; Gwaltney, SR.; Heyden, A.; Hirata, S.; Hsu, C-P.; Kedziora, G.; Khalliulin, RZ.; Klunzinger, P.; Lee, AM.; Lee, MS.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, EI.; Pieniazek, PA.; Rhee, YM.; Ritchie, J.; Rosta, E.; Sherrill, CD.; Simmonett, AC.; Subotnik, JE.; Zhang, W.; Bell, AT.; Chakraborty, AK.; Chipman, DM.; Keil, FJ.; Warshel, A.; Hehre, WJ.; Kong, J.; Krylov, AI.; Gill, PMW.; Head-Gordon, M. Q-Chem, Version 3.0. Q-Chem, Inc.; Pittsburgh, PA: 2006. Jr., R. A. D.III, H. L. W.III, H. F. S.
- [66]. Ponder, JW. TINKER, Software Tools for Molecular Design, Version 4.2. Jun. 2004 The most updated version for the TINKER program can be obtained from J. W. Ponder's World Wide Web site at <http://dasher.wustl.edu/tinker>
- [67]. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. *J. Chem. Phys* 1984;81:3684–3690.

Figure 1. Overall structure of Ca²⁺-bound PAD4 in complex with histone peptide H4 (PDB ID: 2DEY) [24] with the N-terminal domain colored in gray and the catalytic domain colored in red. Ca²⁺ ions and the substrate are shown as black balls and green sticks, respectively.

Scheme 1: starting from the state I.

Scheme 2: starting from the state N.

Figure 2.

Proposed reaction mechanism of protein citrullination by PAD4. The reactants of Scheme 1 and 2 correspond to two probable reactant states (I and N) of PAD4 respectively.

Figure 3. QM/MM partition of the neutral model. Four boundary atoms are colored red and all the other QM atoms are blue or green. The other atoms belong to the MM subsystem. For the ionic model, the partition is the same but the hydrogen of green is bonded to N_δ of His471.

Figure 4.

Active site structure comparison among the initial model of the crystal structure, the ionic (I) and neutral (N) models of the reactant-complex at the end of 35 ps ab initio QM/MM MD simulations. For the initial model of the crystal structure, the Ala645 has been mutated back to Cys while all other coordinates are the same as in the crystal structure 2DEY. The labeled distances between S and C ζ atoms for state I and N are calculated by averaging 1000 snapshots in the last 10 ps simulation.

Figure 5.

The calculated free energy profile for the change from the neutral state to the ionic state, which was determined with B3LYP(6-31G*) QM/MM MD simulations with umbrella sampling. The initial structure employed is the neutral model of the complex, and the reaction coordinate was chosen as the difference between two bond lengths: $RC = d_{HS...S} - d_{HS...N_{\eta 1}}$. A total of 18 umbrella windows have been employed, covering the reaction coordinate $-0.80 \sim 1.70$ Å, so that the total length of ab initio QM/MM MD simulations for this reaction path is 540 ps.

Figure 6.

Illustration of the structures along the reaction path from the neutral state to the ionic state as in Fig. 5. The snapshots correspond to the RC values of 0.6 Å and 1.5 Å, respectively.

Figure 7.

The calculated free energy profile for the deimination reaction catalyzed by PAD4. In the reactant complex, both Cys645 and His471 are neutral. The overall deimination reaction consists of two reaction steps: the formation of a tetrahedral intermediate and the cleavage of $C_{\zeta} - N_{\eta 1}$ bond. In the first step, the reaction coordinate was chosen as the combination of four bonds: $RC = -d_{HS \cdots N_{\eta 1}} - d_{S \cdots C_{\zeta}} - 2d_{H \cdots N_{\eta 1}}$, while in the second step, the breaking $C_{\zeta} - N_{\eta 1}$ bond was chosen as the reaction coordinate. The two $d_{H \cdots N_{\eta 1}}$ refer to as the two initial $H - N_{\eta 1}$ bonds in the amide group. It is found that those bonds would break if they are not included into the reaction coordinate. A total of 31 simulation windows with harmonic potential force constants of $20 \sim 60 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ have been employed. The total length of B3LYP(6-31G*) QM/MM MD simulations to determine this free energy profile is 930 ps.

Figure 8. Illustration of the structures along the deimination path. They correspond to the reactant, two configurations of the transition state I and the tetrahedral, transition state II, thionurinium intermediate.

Figure 9.

Time dependence of the two $H-N_{\eta 1}$ bonds for 50 ps QM/MM MD simulation for the transition state window. H_{11} and H_{12} are two protons hydrogen bonded with His471 and Asp473 respectively.

Figure 10.

The calculated free energy profile for the ammonia diffusion. The reaction coordinate is chosen as the distance between the geometry centers of the ammonia molecule and the residues around the active site: His471, Asp473, Asn588, His644 and Cys645. A total of 27 windows were employed, covering the distance from 3.0 to 16.0 Å, with a harmonic force constant of 10 kcal·mol⁻¹ · Å⁻². For each window, MD simulation consisted of 500 ps equilibration and 1000 ps sampling.

Figure 11.

The left panel illustrates the NH_3 diffusion path. Each NH_3 molecule is labeled by the corresponding reaction coordinate. The substrate is shown by pink stick, while the NH_3 is colored green. The path is lighted with a red line. The right panel is the comparison of the structures before and after the leaving of NH_3 . The structure with RC of 4.0 Å is colored yellow, while that of RC 13.0 Å is in blue.

Table 1

List of the key geometric parameters of the reactant, the transition states, and the intermediates from B3LYP(6–31G*) QM/MM MD simulations

Distance (Å)	Reactant	TS I	Int. I	TS II	Int. II
S(Cys645)-C _α (Sub)	3.72±0.15	2.76±0.22	1.93±0.06	1.83±0.04	1.79±0.04
C _α (Sub)-N _ε (Sub)	1.34±0.02	1.34±0.03	1.40±0.03	1.35±0.02	1.34±0.02
C _α (Sub)-N _{η1} (Sub)	1.34±0.02	1.44±0.03	1.58±0.06	2.10±0.01	4.40±0.01
C _α (Sub)-N _{η2} (Sub)	1.35±0.02	1.35±0.04	1.42±0.03	1.36±0.02	1.31±0.02
N _ε (Sub)-O _{δ2} (Asp350)	2.77±0.09	2.87±0.18	2.83±0.11	2.80±0.10	3.00±0.37
N _{η1} (Sub)-N _δ (His471)	3.34±0.16	2.80±0.11	2.90±0.11	3.04±0.13	5.98±0.28
N _{η1} (Sub)-O _{δ2} (Asp473)	2.82±0.11	2.72±0.12	2.79±0.10	2.89±0.12	4.49±0.90
N _{η2} (Sub)-O _{δ1} (Asp350)	2.76±0.09	3.47±0.50	3.18±0.18	2.91±0.21	3.80±0.31
N _{η2} (Sub)-O _{δ1} (Asp473)	3.15±0.17	2.95±0.13	3.01±0.25	3.04±0.14	3.36±0.20