Theor Chem Acc (2012) 131:1293
DOI 10.1007/s00214-012-1293-9

REGULAR ARTICLE

Semimicroscopic investigation of active site pK_a values in peptidylarginine deiminase 4

Ruthanne S. McCoy · Sonja B. Braun-Sand

Received: 8 February 2012 / Accepted: 20 October 2012

Abstract Peptidylarginine deiminase 4 (PAD4), also known as protein arginine deiminase 4, performs a post-translational deimination that converts arginine to citrulline. The dysregulation of PAD4 has been implicated in a number of diseases, including rheumatoid arthritis (RA) and cancer. This makes PAD4 an important therapeutic target. To develop small-molecule inhibitors as potential treatments, it is advantageous if the catalytic mechanism is well understood. The protonation states of the active site residues, which have long been under controversy, have a direct impact on the catalytic mechanism. Two competing mechanisms are under investigation in the current literature. The first is a reverse protonation mechanism that depends on the active site histidine and cysteine existing as an ion pair. The second is a substrate-assisted mechanism that depends on the active site histidine and cysteine being neutral. This study uses the semimicroscopic protein dipoles Langevin dipoles (PDLD/S) linear response approximation method in the MOLARIS software package to calculate the change in solvation energy of moving the residue from water to the protein interior, and then using that information to assess the protonation states of the active site residues of PAD4. Results from these calculations suggest that in the enzyme–substrate complex of PAD4, the cysteine and histidine are protonated and deprotonated, respectively, and are therefore both neutral, analogous to the proposed protonation states of the active site residues in the Michaelis complex in the substrate-assisted mechanism.

Keywords Protein arginine deiminase 4 · Peptidylarginine deiminase 4 · PAD4 · PADIV · Rheumatoid arthritis · Substrate-assisted mechanism

1 Introduction

Many physiological roles of PAD4 have been elucidated. In HL-60 granulocytes, PAD4 is localized in the nucleus and targets its catalytic activity on histones H2A, H3, and H4 [1–5]. Neutrophil extracellular traps, or NETs, form upon histone citrullination in HL-60 granulocytes. NETs are highly decondensed chromatin structures that are thought to form as an innate immune response to a bacterial infection [6]. PAD4 is also involved in the p53 tumor suppression pathway [7, 8]. Inhibition of PAD4 has been shown to increase the expression of the genes p21, C1P1, and WAF1, which are all target genes of p53 and are responsible for regulating the cell cycle and apoptosis [8]. Furthermore, protein–protein interaction studies have shown that p53 interacts directly with PAD4, to take PAD4 to the p21 promoter, where it citrullinates histones to repress p21 expression [8].

PAD4 has been implicated in a number of diseases, most notably rheumatoid arthritis (RA). Hypercitrullination may be a factor in the progression of RA [9–12]. Autoantibodies, such as rheumatoid factor, antiperinuclear factor, and antikeratin autoantibody, target citrullinated proteins and may be responsible for the joint damage associated with RA [10]. In addition, citrullinated proteins are found in much higher concentrations in the synovial membranes

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1293-9) contains supplementary material, which is available to authorized users.

R. S. McCoy · S. B. Braun-Sand (✉)
Department of Chemistry and Biochemistry, University of Colorado Colorado Springs, 1420 Austin Bluffs Pkwy, Colorado Springs, CO 80918, USA
e-mail: sbraunsa@uccs.edu

Published online: 09 November 2012

of RA patients than in control subjects [11]. Genetic abnormalities may increase the risk of developing RA. In one study, eight single-nucleotide polymorphisms (SNPs), four of which were found in exons, were associated with an increased risk of rheumatoid arthritis [12]. These results were found in Japanese patients, and further investigation is needed to assess whether the same correlation is found in other populations [9].

Human peptidyl arginine deiminase 4 (PAD4) is a member of the guanidino group-modifying enzyme superfamily (GMSF) [13]. In addition to PAD4, members include arginine deiminase (ADI) [13–24], dimethylarginine dimethylaminohydrolase (DDAH) [25], and amidinotransferase (AT) [13, 26–30]. Though these enzymes have a low sequence similarity (approximately 8–14 %), they share a common active site fold and basic catalytic motif (Cys–His–Asp/Glu) [13]. PAD4 catalyzes the calcium-dependent, post-translational modification of arginine residues to citrulline (Fig. 1) through modification of the guanidinium group [1, 9, 31–33]. The quaternary structure of PAD4 is a dimer made of two identical monomers [31]. Each PAD4 monomer contains five Ca^{2+}-binding sites, and structural data from Arita et al. [31] indicate that the active site cleft is formed upon binding of Ca^{2+} ions, which causes conformational changes in the protein. Two Ca^{2+} ions are located in the C-terminal domain (also the catalytic domain), while the other three are located in the N-terminal domain. The N-terminal domain is composed of two immunoglobulin (Ig)-like subdomains, and three Ca^{2+} binding sites are located near the surface of the second Ig-like subdomain [31]. X-ray crystallographic studies by Arita et al. show that in the absence of Ca^{2+}, the second subdomain is disordered, but in the presence of the Ca^{2+}, it forms an ordered α-helix at the protein surface. These conformational changes may be important for a protein–protein interaction or for a Ca^{2+}-mediated regulatory mechanism [31]. Overall, the N-terminal domain is far from the active site and is not believed to directly affect the catalytic function [34, 35]. For this reason, the N-terminal domain was removed from recent theoretical studies [34, 36]. The C-terminal, catalytic domain consists of five ββαβ motifs forming a pseudo-fivefold structure called the α/β propeller, and at the center is located the catalytic triad (Cys–His–Asp) [31]. The two Ca^{2+} in the C-terminal domain are located at the bottom of the active site and are believed to induce activation of PAD4 [31].

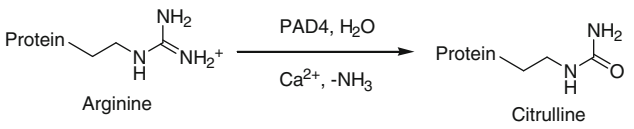

Fig. 1 Reaction catalyzed by PAD4, showing an arginine sidechain of the protein substrate

The active site contains four residues that are essential for catalysis: Asp350, His471, Asp473, and Cys645 as indicated by mutagenesis studies [31, 37]. His471 and Cys645 are directly involved in the mechanism, while Asp473 and Asp350 coordinate with the substrate via hydrogen bond formation to hold it in place during catalysis [34, 37]. It is generally thought that PAD4, like other members of the GMSF family, uses a two-stage catalytic mechanism [13, 16, 17, 33, 37, 38]. The first step is the deimination reaction, followed by a hydrolysis. The deimination reaction involves a nucleophilic attack on the substrate guanidinium carbon atom (C_ζ) performed by Cys645 leading to cleavage of the C_ζ–$N_{\eta 1}$ bond to release ammonia, as shown in Fig. 2. In the second stage, a water molecule replaces the ammonia and performs hydrolysis of the thiouronium intermediate, forming the product citrulline and regenerating the enzyme active site.

Two catalytic mechanisms have been proposed to carry out this two-stage reaction: the "reverse protonation mechanism" (Fig. 2a) [37], and the "substrate-assisted mechanism" (Fig. 2b) [31]. As can be seen in Fig. 2, which shows the deimination step only, both proposals reach the same intermediate even though the starting structures of the reactions differ. The reverse protonation mechanism involves a nucleophilic attack of the Cys645 thiolate anion on the guanidinium group of peptidyl–arginine followed by the donation of a proton from His471 to form a tetrahedral intermediate with ammonia as the leaving group (see also Supporting Information for Figure SI-1 depicting this mechanism) [37, 39]. This mechanism depends on His471 and Cys645 existing as an ion pair in the enzyme–substrate (ES) complex. In this proposed mechanism, there is a high concentration of charge in the active site: two negatively charged aspartates, a negatively charged cysteine, and a histidine and substrate that are positively charged. Though this mechanism proposes five charges concentrated in the active site, there has been a previous study examining other catalytic sites that also have high charge concentrations [40]. This mechanism was proposed for PAD4 because the structures of GMSF proteins (including PAD4) have no base close enough to the active site cysteine to deprotonate it, and because thiols are poor nucleophiles, it was assumed that the cysteine must already be deprotonated in the Michaelis complex, similar to cysteine proteases [25, 41, 42].

The alternate mechanism (Fig. 2b), originally proposed by Arita et al. [31] and further explored by Ke et al. [34, 36], is referred to as the substrate-assisted mechanism and begins with His471 and Cys645 both neutral (see Supporting Information, Figure SI-2, for a depiction of this mechanism). In this mechanism, the deimination reaction involves a concerted proton abstraction from Cys645 by

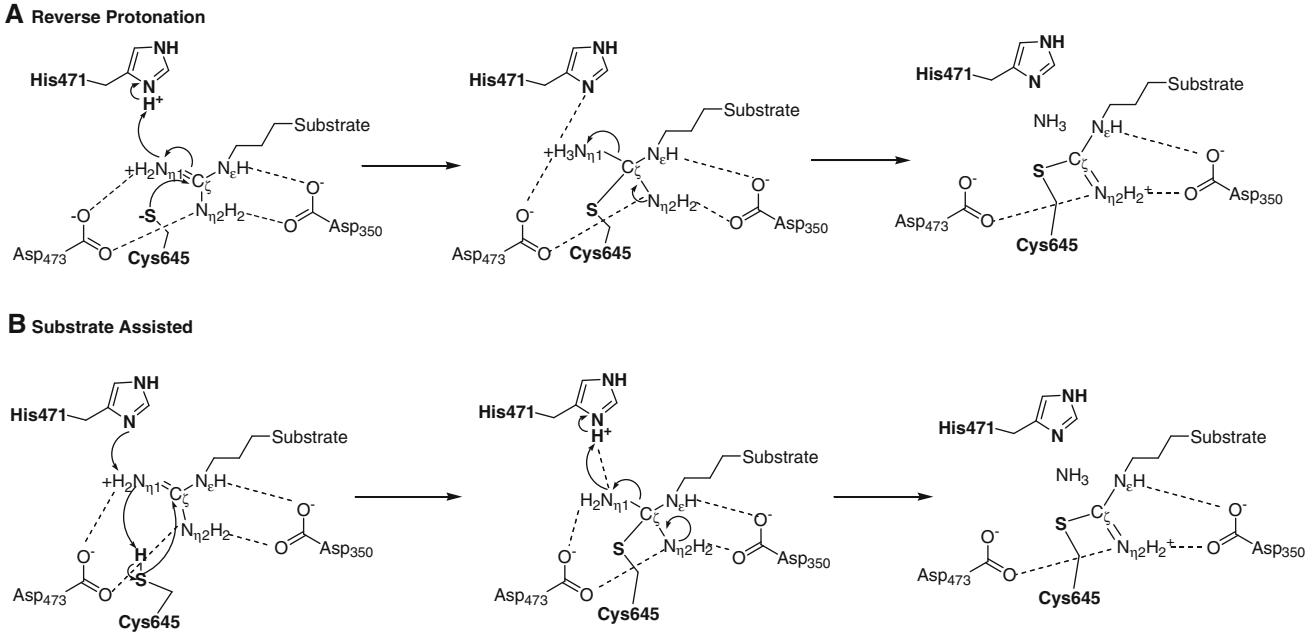

Fig. 2 **a** This figure shows the steps in the deimination reaction starting from a Cys–His ion pair. **b** This figure shows the steps in the deimination reaction starting from neutral Cys and His. The thiouronium intermediate (far *right*, *top*, and *bottom*) is the same for both potential starting structures

the substrate $N_{\eta 1}$ of the guanidinium group with a nucleophilic attack of the thiolate at the C_ζ of the guanidinium group. This is followed by cleavage of the C_ζ–$N_{\eta 1}$ bond to give ammonia and the thiouronium intermediate. This mechanism was proposed because His471 and Cys645 are found to be $\sim$6 Å apart in PAD4 crystal structures, which is thought to be too far for an ion pair to be stabilized [34]. The active site histidine to cysteine distance is even larger in crystal structures of arginine deiminase (ADI) [see, for example, Galkin et al. [15], which indicates the Cys–His distance in *Pseudomonas aeruginosa* ADI is $\sim$7 Å]. Others have postulated that ADI and DDAH have their active site cysteine residues protonated in the absence of substrate, and only lose the proton upon substrate binding [21, 38]. A mutation of the active site histidine to a glycine in *P. aeruginosa* DDAH found that it did not eliminate cysteine nucleophilicity, an argument against a preformed ion pair [38]. The optimum pH of GMSF family members spans a wide range, from pH less than 5.5 for *P. aeruginosa* ADI [17, 19] to pH 7.6 for human PAD4 [33]. Due to the low optimum pH for ADI, it is perhaps more likely to use the substrate-assisted mechanism than PAD4; however, even if the reverse protonation mechanism is correct for PAD4, only about 15 % of active enzyme would exist at the optimum pH of 7.6 (see Figure 8 in Ref. [34]).

Rastogi and Girvin have shown that internal ionizable residues can have their pK_a values altered during a cycle of function, as the residue can be located in different microenvironments during the cycle [43]. The study reported here is important because it gives insight into the most probable protonation states of active site residues of the human PAD4 active site both before substrate binds and after, which may open a window into the mechanism. The results of this study show, in some instances, large differences in pK_a values in the protein active site compared to solution values. A recent experimental work by Isom et al. [44] has shown that it is possible to observe large changes in pK_a values in a protein interior compared to typical pK_a values in water. This work reported the engineering of 25 variants of staphylococcal nuclease (SNase) and demonstrated that 19 of the 25 variants had an interior lysine residue with a depressed pK_a value, in one instance as low as 5.3 (from a solution value of 10.4), while other variants displayed little to no change in the lysine pK_a value. The origin of the shifts of these pK_a values upon moving to the protein interior is explained in relation to the differing microenvironments of the ionizable groups, similar to explanations given previously for the c subunit of ATP synthase [43]. Isom and coworkers reason that in a highly polar or polarizable microenvironment, the pK_a value will differ little relative to water. However, in a less polar or polarizable microenvironment (such as a hydrophobic protein interior), the pK_a value can change significantly, and greater changes in pK_a values are observed. In less polarizable microenvironments, acidic sidechains will tend to have higher pK_a values than in water [45–47], while basic sidechains will tend to have lower pK_a values than in water [48–50]. Similarly, Rastogi and Girvin [43] reported evidence that when Asp61 in subunit c of ATP synthase was protonated, it was likely buried in a packing interface

between the N- and C-terminal helices of subunit c. In contrast, when the Asp61 was deprotonated, it was likely on an exposed helical face.

Computational studies examining the mechanism of PAD4 have been previously published [34, 36, 39] and reached differing conclusions. The proposed substrate-assisted mechanism and reverse protonation mechanism differ only in the first step of the reaction (the deimination step) and lead to the same intermediate (see Fig. 2 here or Figure 2 in Ref. [34]). Born–Oppenheimer ab initio QM/MM studies of PAD4 indicated that in the Michaelis complex, the active sites Cys645 and His471 are both neutral prior to the reaction and that the Cys645 is deprotonated by the substrate guanidinium group in concert with the nucleophilic attack by the thiolate at the C_ζ position of the guanidinium group, thus forming a tetrahedral intermediate [34]. This concerted first step was found to be the rate-determining step of the reaction [34, 36], with a barrier of 20.9 kcal/mol, which agrees well with experimental kinetic studies which found a k_{cat} of 6.55 s^{-1}, corresponding to a free energy barrier of 17.0 kcal/mol [33]. Further work by Ke et al. [36] explored the hydrolysis of the thiouronium intermediate by the same method. Results indicated that the barrier for the hydrolysis was 16.5 kcal/mol, which confirmed that the deimination step was rate-determining.

In contrast, an earlier density functional study [39] used a model of the enzyme active site which included a portion of the sidechains of His471, Asp473, Asp350, Cys645, and a substrate model of N-ethylguanidinium. This study carefully explored whether the rate-determining step was the deimination reaction or the hydrolysis reaction, and also compared gas phase energetics to solvated energetics of reaction. This work found the rate-determining step in the gas phase to be the ammonia release, with a barrier of 10.8 kcal/mol. Use of a CPCM solvent with a dielectric constant of 4 to model the protein environment found that the rate-determining step increased by approximately 3.2 kcal/mol. This work examined both stages of the catalytic reaction to determine whether the deimination or the hydrolysis was rate-determining, but only modeled one beginning scenario, the case with histidine positively charged and cysteine negatively charged. Comparison of the reverse protonation mechanism with the substrate-assisted mechanism was not the aim of this study; thus, it is not possible to determine which mechanism is energetically more favorable. In addition, use of a dielectric constant to model a protein environment ignores the electrostatic contributions of the surrounding environment, and its limitations are well documented (see for example [51–53]).

All of these aforementioned studies are very informative, but a study such as the one reported here is necessary, as none of the previous studies examined the pK_a^p values of the active site residues either with or without substrate bound, which arguably are important for a fuller understanding of the mechanism of the enzyme. A previous experimental study found active site residues in the presence of benzoyl-L-arginine ethyl ester (BAEE) for human PAD4 to have pK_a values of 8.2 and 7.3 [37]. In solvent, the sidechains of cysteine and histidine have pK_a values of 8.3 and 6.0, and therefore, it is not unreasonable to assign the pK_a^p value of 8.3 to the cysteine and the pK_a^p value of 7.3 to the histidine. In fact, Knuckley et al. [37] postulate that the values of 7.3 and 8.2 likely correspond to His471 and Cys645, respectively, before substrate binds. Considering the highly polar/polarizable environment of the PAD4 active site, it could be reasoned that the pK_a^p values would differ little relative to water [44]. At the PAD4 optimum pH of 7.6, if the experimental pK_a^p value of 7.3 is assigned to his and the pK_a^p of 8.2 is assigned to cys, the majority of his would be deprotonated and the majority of cys would be protonated, as in the substrate-assisted mechanism. However, upon further investigation of the kinetics, the authors found evidence that supports that inactivators (iodoacetamide and 2-chloroacetamidine) bind preferentially to the thiolate form of the enzyme. The authors write: "The fact that the iodoacetamide and 2-chloroacetamidine inactivation kinetics yield similar pK_a values for Cys645 is inconsistent with a pure substrate-assisted mechanism of thiol deprotonation" [37]. Arguably, definitive assignment of experimental pK_a^p values to particular residues is difficult [37].

Because of the difficulty in assigning pK_a^p values experimentally, a study such as the one reported here is helpful in understanding the character of the active site of PAD4 with and without a positively charged substrate bound. This study provides computational details of the stabilities of the active site residues under many different permutations of the ionization states of these residues to compare the stability of various active site protonation states. The difference in solvation energies of these residues in water and in protein were calculated to assess the pK_a^p values of residues in the active site, which have a direct impact on the catalytic mechanism utilized by the enzyme. The pK_a^p values reported here are not intended to correspond to directly observable pK_a^p values; rather, they should be interpreted as a tool to help understand the population of various microstates of the PAD4 active site [54].

2 Methods

2.1 Permutations of active site conditions for calculations to determine pK_a^p values

Human PAD4 crystal structures 1WDA [31], containing the positively charged synthetic substrate benzoyl-L-arginine

amide (BAG) (Fig. 3) and 1WD9 [31], an enzyme structure without ligand bound, were obtained from the Brookhaven protein data bank [55, 56]. Both crystal structures were mutants of human PAD4 containing an alanine residue in place of the active site cysteine. The alanine sidechain was computationally mutated back to the cysteine sidechain to obtain the catalytically competent sequence of the enzyme.

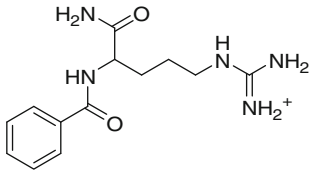

Fig. 3 Structure of synthetic substrate benzoyl-L-arginine amide (BAG)

Four combinations of protonated and deprotonated active site histidine and cysteine sidechains were used to ascertain pK_a^p values for each ionizable active site residue (Asp350, Asp473, His471, Cys645) and the substrate (BAG) when present. The crystal structures with (1WDA) [55] and without (1WD9) [56] substrate were used as starting points for the calculations. The permutations of active site residues that were examined computationally are shown in Fig. 4. The calculations of these permutations were performed both with and without substrate and include: cys ionized/his neutral (Cys^-/His^0), cys neutral/his ionized (Cys^0/His^+), cys ionized/his ionized (Cys^-/His^+), cys neutral/his neutral (Cys^0/His^0).

In all calculations, the active site aspartic acid residues and the substrate were always ionized (if substrate was present). This was due to the very low water pK_a value of aspartic acid and the high water pK_a value of BAG. There

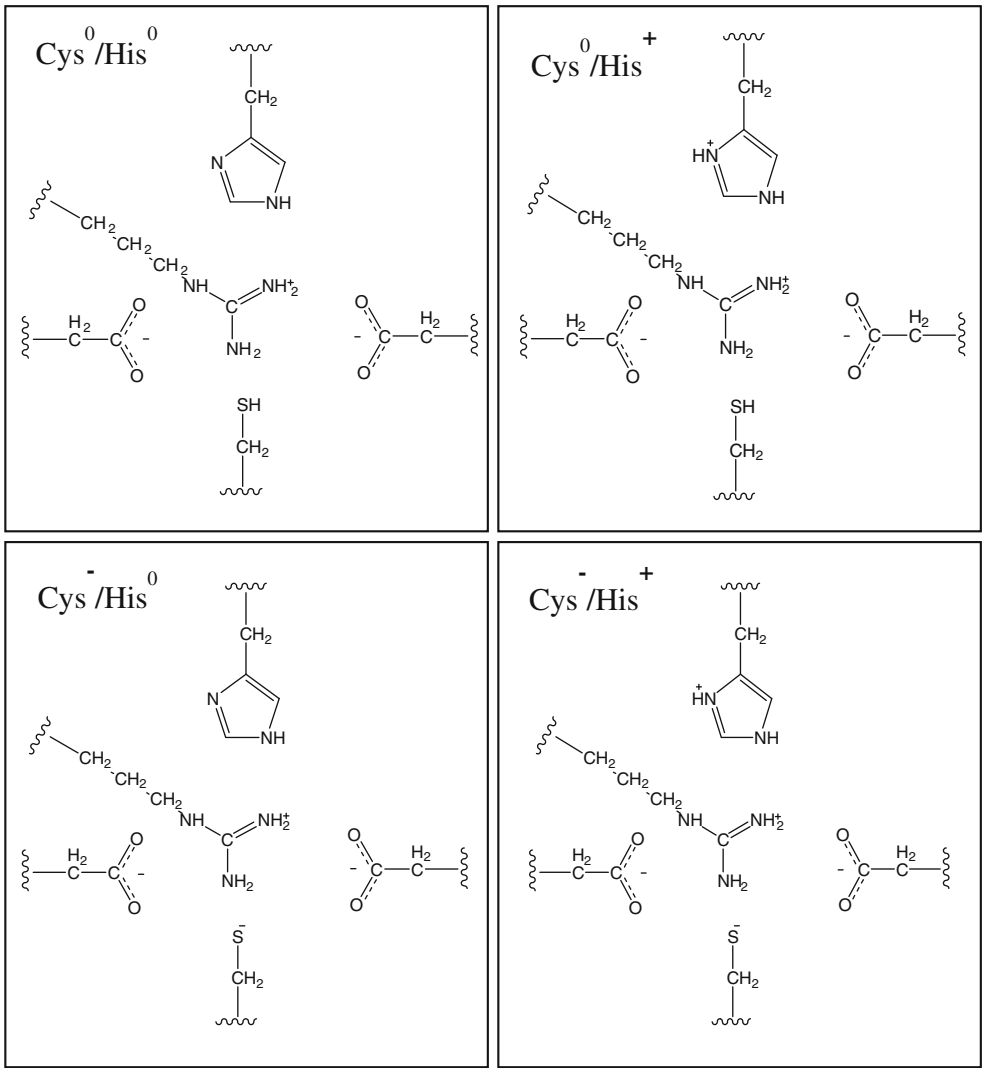

Fig. 4 Schematic drawing of the four permutations of the ionizable active site residues that were examined in this study

is some experimental evidence that suggests these aspartic acids must be negatively charged for PAD4 to bind [37], indicating an electrostatic interaction between the side chains of the aspartic acid residues and the guanidinium group of BAG. BAG was considered to have a solution pK_a value equivalent to arginine's sidechain solution pK_a value, due to the presence of the guanidinium group. The three calcium atoms in the N-terminal domain that have been determined to have little effect on catalysis [34], and are in a different domain than the active site [31], were neutralized. The two calcium atoms in the catalytic domain were ionized, and surrounding acidic residues were ionized to balance the charge in this region. These residues were Glu411, Glu353, Glu351, and Asp369. Three sulfate molecules, present due to the crystallization process but not necessary for catalysis and far from the active site, were assigned a charge of zero. In total, 364 pK_a^p values were calculated, and each reported number is an average of thirteen trials, reported with the standard deviation of the trials.

2.2 Software and model system

SPARTAN (version 08) [57, 58] was used to calculate the partial charges on the BAG using the restricted Hartree–Fock (RHF) calculation method with a 3-21G basis set, followed by a density functional RB3LYP method with a 6-31+G* basis set. These charges were used as input (see Table SI-1 in the Electronic Supporting Information for actual partial charges used) along with the PAD4 crystallographic data for the software package MOLARIS, version 9.05 [59]. MOLARIS was used to perform molecular dynamics (MD) simulations on the model system, and the protein dipoles Langevin dipoles semimacroscopic linear response approximation (PDLD/S-LRA) method in MOLARIS was used to calculate the pK_a^p of each of the residues of interest [60, 61].

The PDLD model explicitly considers the protein/solvent system with all of its electrostatic components. The PDLD method breaks the protein/solvent system up into four regions. Region I contains the charged group of interest (for which pK_a^p will be calculated), region II contains the protein atoms found within a specified radius from the center of region I (22 Å in this study), region III represents the water molecules by a Langevin grid and has a specified radius from the center of region I (22 Å in this study), region IVa contains the rest of the protein atoms outside region II, and Region IVb is the bulk solvent beyond the specified radius. Regions I, II, and III have the electrostatic effects treated explicitly, while the electrostatic effects in regions IVa and IVb are treated by a macroscopic continuum formulation (for more details and a figure showing the regions, please see Ref. [61]). The effective PDLD potential of a charged group is represented by:

$$\Delta V_{\mathrm{PDLD}} = \Delta V^p_{q\mu} + \Delta V^p_{q\alpha} + \Delta V^p_{qq} + \Delta V^p_{qw} + \Delta G^p_{\mathrm{bulk}} \qquad (1)$$

where $\Delta V^p_{q\mu}$ gives the interaction between the charge and the protein permanent dipoles, $\Delta V^p_{q\alpha}$ gives the interaction between the charge and the protein-induced dipoles, ΔV^p_{qq} gives the interaction between the charge and other ionized groups, ΔV^p_{qw} gives the interaction between the charge and the Langevin dipoles (which represent the average polarization of water molecules in and around the protein), and $\Delta G^p_{\mathrm{bulk}}$ is the solvation energy of the bulk solvent, which surrounds the region of explicit water molecules.

A strength of the PDLD approach is the treatment of protein structural relaxation upon charge formation within the linear response approximation (LRA) framework. The PDLD results are averaged over MD-generated protein configurations both with the residue of interest being neutral and with the residue of interest being ionized, as shown below:

$$\Delta G_{\mathrm{PDLD}} = \frac{1}{2}\left[\langle \Delta V_{\mathrm{PDLD}} \rangle_{r^p_{q=0}} + \langle \Delta V_{\mathrm{PDLD}} \rangle_{r^p_{q=\bar{q}}}\right]. \qquad (2)$$

It should be noted that q can be equal to 0, when the residue is not ionized, or q can be equal to $\bar{q}$, which is typically ± 1. The bar over the q does not represent an average; rather, it denotes that the residue is in the ionized state. The $\langle\rangle_{r^p}$ denotes an average over protein configurations with the assigned q, and ΔV_{PDLD} is defined in Eq. 1. The PDLD model has large microscopic contributions. To obtain more stable results, the microscopic contributions can be scaled in a consistent fashion using the semimicroscopic PDLD model (PDLD/S) [60, 62]. The PDLD/S model represents the contributions that are not included explicitly in the model by assigning the protein a "dielectric constant," ϵ_p. The ϵ_p should be viewed primarily as a scaling factor, and not an actual protein dielectric constant [52, 61]. MOLARIS assigns dielectric constants based on protein-induced dipoles; this constant falls between 2 and 40 in MOLARIS; a value of 4 was used for this study as it was found to be the optimal value for the PDLD/S-LRA model [52, 53]. The PDLD/S effective potential is represented by:

$$\Delta V^{w\to p}_{\mathrm{pdld}/s} = -\left[\Delta G^w_{qw} + \left(\Delta G^p_{qw}(q=\bar{q}) - \Delta G^p_{qw}(q=0)\right)\right]\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right) + \left(\Delta V^p_{qq}(q=\bar{q}) + \Delta V^p_{q\mu}(q=\bar{q})\right)\frac{1}{\varepsilon_p} \qquad (3)$$

where ΔG^w_{qw} is the self-energy of the charge in water, ΔG^p_{qw} is the change in solvation energy of the protein with and

without the charged group, and ΔV^p_{qq} and $\Delta V^p_{q\mu}$ are the same as in Eq. 1. Again, the LRA framework ensures that protein reorganization is considered by taking into account the relaxed structures in both the ionized and neutral states of the relevant charge [61], and the PDLD/S-LRA free energy is evaluated in the same fashion as Eq. 2, but substituting the PDLD/S effective potential (Eq. 3) for the PDLD effective potential, using:

$$\Delta G_{\mathrm{PDLD/S}} = \frac{1}{2}\left[\left\langle \Delta V_{\mathrm{PDLD/S}} \right\rangle_{r^p_{q=0}} + \left\langle \Delta V_{\mathrm{PDLD/S}} \right\rangle_{r^p_{q=\bar{q}}}\right]. \quad (4)$$

The all-atom MD simulations were used to generate protein conformations [61]. In the absence of BAG (PDB ID 1WD9), multiple MD simulations were run for each of the permutations shown in Fig. 4, having as their center one of the four residues for which the pK^p_a values were to be calculated. In the presence of BAG (PDB ID 1WDA), multiple MD simulations were run for each permutation, with either the BAG or one of the four residues for which the pK^p_a values were to be calculated at the center. As an example, for the Cys^-/His^+ permutation in the presence of BAG, five MD simulations were run, having at the center a residue for which the pK^p_a was to be calculated: Asp350, Asp473, His471, Cys645, or BAG. The pK^p_a value of a residue can only be calculated if it is charged in that permutation, so for the Cys^0/His^+ permutation in the presence of BAG, four MD simulations were run having as their center Asp350, Asp473, His471, or BAG.

The MD simulations used a timestep of 1 fs and were run for 120 ps at 100 K, 120 ps at 200 K, and at least 390 ps at 300 K. Longer simulation times are not required for equilibration and water penetration because the PDLD/S-LRA simulations use a dielectric constant that reflects the missing water penetration, as described above. This is a fact that has been established by the PDLD/S-LRA validation and has been shown in very challenging case like internal groups in staphylococcal nuclease, cytochrome c oxidase, and others [63–68].

Thirteen conformers were abstracted from the MD simulations performed at 300 K. The conformers were snapshots taken every 30 ps, starting from the previous conformer. These thirteen conformations were generated with the center of the MD simulation being the residue for which the pK^p_a value was to be calculated. Each conformer was then used as a starting structure in the PDLD/S-LRA solvation calculations used in this study to determine the pK^p_a values.

The PDLD/S-LRA solvation method performs calculations based on a thermodynamic cycle (shown in Fig. 5) developed by Warshel et al. [61, 69]. This thermodynamic cycle considers the free energy contributions associated with ionizing an acidic residue in a protein. A mathematical representation of this cycle is shown in Eq. 5 [52, 53]:

$$\begin{aligned}\Delta G^p\left(\mathrm{AH}_p \rightarrow A^-_p + H^+_w\right) &= \Delta G^w\left(\mathrm{AH}_w \rightarrow A^-_w + H^+_w\right) \\ &\quad + \Delta G^{w\rightarrow p}_{\mathrm{sol}}(A^-) - \Delta G^{w\rightarrow p}_{\mathrm{sol}}(\mathrm{AH})\end{aligned} \quad (5)$$

Designations p and w stand for protein and water, respectively. $\Delta G^{w\rightarrow p}_{\mathrm{sol}}$ provides the change in Gibbs free energy of the acidic (AH) or basic (A^-) species as it moves from water to the protein active site. This thermodynamic cycle can be visualized pictorially in Fig. 5.

To evaluate the $\Delta G^{w\rightarrow p}_{\mathrm{sol}}$ term of Eq. 5, the self-energy of ionizing this group when all other ionizable groups, such as other residues in the active site, are uncharged is first considered, and then the effect of charging the other groups to their ionization state is considered. This term can then be expressed as

$$\begin{aligned}\left(\Delta G^{w\rightarrow p}_{\mathrm{sol}}\right)_i &= \left(\Delta G^p_{\mathrm{self}} - \Delta G^w_{\mathrm{self}}\right)_i + \sum_{i\neq j}\Delta G^p_{ij} \\ &= \left(\Delta G^p_{q\mu} + \Delta G^p_{q\alpha} + \Delta G^p_{qw} - \Delta G^w_{\mathrm{self}}\right)_i + \sum_{i\neq j}\Delta G^p_{ij}\end{aligned} \quad (6)$$

where ΔG_{self} is the self-energy associated with charging group "i" in its environment. For proteins, this can be broken down into the interaction of the charge on group "i" with the surrounding permanent dipoles ($\Delta G_{q\mu}$), the surrounding induced dipoles ($\Delta G_{q\alpha}$), and the water molecules in and surrounding the protein ΔG_{qw}. This is described more fully elsewhere [52, 61, 71].

Equation 5 can also be written in terms of the pK_a value of the ith ionizable residue in the protein, $pK^p_{a,i}$, as

$$pK^p_{a,i} = pK^w_{a,i} - \frac{\bar{q}_i}{2.3RT}\Delta\Delta G^{w\rightarrow p}_{\mathrm{sol}}(\mathrm{AH}_i \rightarrow A^-_i) \quad (7)$$

with the $\Delta\Delta G^{w\rightarrow p}_{\mathrm{sol}}$ consisting of the last two terms of Eq. 5, and $pK^w_{a,i}$ is the pK_a value of the ith residue in water. The $\bar{q}_i$ term is the charge on the ionized form of the residue. It should be noted that there are two possibilities for the charge, q_i. The charge can have $q_i = \bar{q}_i$, which corresponds to the charged form of the residue (-1 or $+1$), or $q_i = 0$, which corresponds to the uncharged form of the residue [52, 71]. It is possible for the acid form of a residue to be either neutral (such as for cysteine, Eq. 8) or positively charged (such as for histidine, Eq. 9). Thus, $\bar{q}_i$ will be negative (corresponding to the charge on the conjugate base),

$$\bar{q}_i = -1(q(\mathrm{AH}) = 0,\ q(A-) = -1) \quad (8)$$

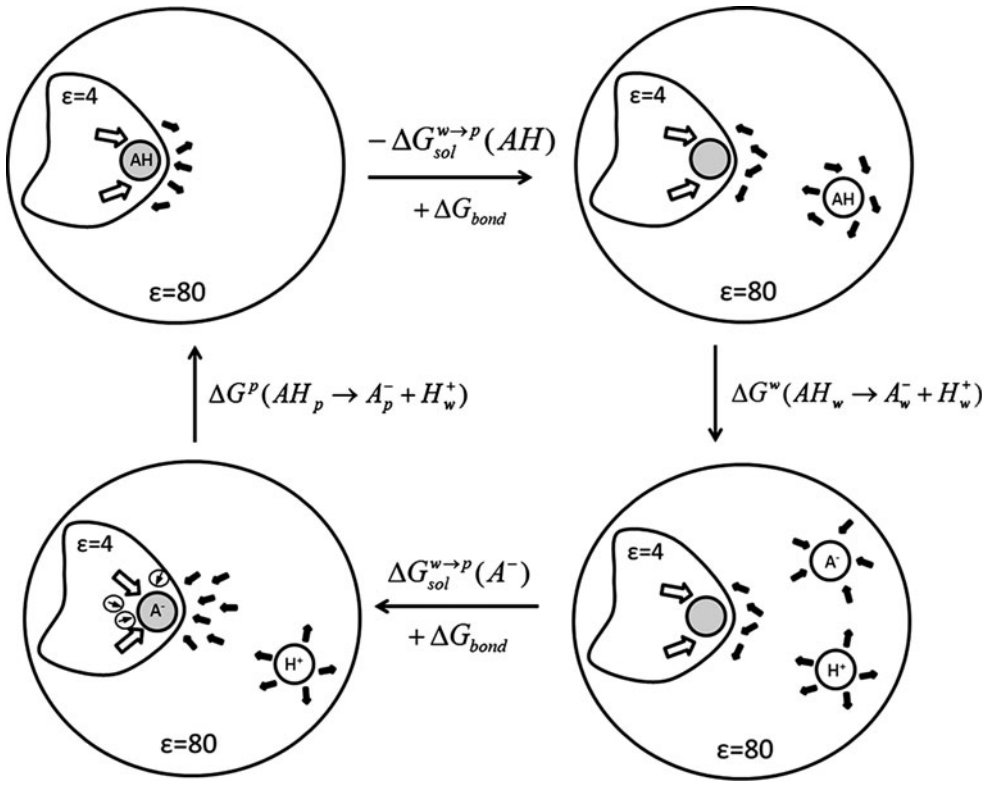

Fig. 5 Pictorial description of the thermodynamic cycle used by MOLARIS software to estimate the ionization energy of an acidic group in a protein active site. The figure describes a fully microscopic cycle and lists the relevant free energy contributions. The subscripts "*p*" and "*w*" designate, respectively, protein and water environments. ΔG_{sol} represents the different solvation free energies, and ΔG_{bond} designates the free energy of breaking the covalent bond between the conjugate acid and the proton. The protein permanent dipoles are represented by *large open arrows*, while the induced dipoles are represented by *small solid arrows* enclosed in *circles*. The dipoles of water molecules are designated by *solid black arrows*. The active site is depicted by a *gray filled circle* within the larger protein (*irregular shape*). To calculate the actual free energies, LRA averaging is performed on the relevant configurations [70]

or positive (corresponding to the charge on the acid),

$$\bar{q}_i = +1(q(\text{AH}) = +1,\ q(A-) = 0). \tag{9}$$

Equation 7 makes possible the calculation of a pK_a value of a residue in the protein active site by evaluating the change in solvation energy for moving the charged group from water to the active site. The $\Delta\Delta G_{\text{sol}}^{w\to p}$ value is extracted from the PDLD/S-LRA solvation calculations and used in Eq. 7 to calculate pK_a^p values. The aim of this investigation is to elucidate the pK_a^p values in PAD4 to give a better understanding of the active site, which may give better insight into the mechanism of PAD4. The computational protocol employed here has been shown to be effective in descriptions of pK_a^p values for other enzymes such as cytochrome c oxidase, SNase, and others [51, 52, 61, 64, 72]. The method used here is simpler and often more reliable than evaluation of the absolute pK_a value by determining the gas phase proton affinity and the solvation of A^- and H_3O^+ [73–76].

3 Results

The average $\Delta\Delta G_{\text{sol}}^{w\to p}$ obtained from the PDLD/S-LRA calculations, the pK_a^p values calculated using Eq. 7, and $[A^-]$:[HA] ratios calculated from the Henderson–Hasselbalch equation are reported in Table 1. The pK_a^p values used in the calculations of microstate populations are italicized. Table 1 includes results for PAD4 only in the absence of BAG using PDB ID 1WD9. The label "Center Residue" corresponds to one of the four active site residues for which pK_a^p values were calculated, and each residue is shown in a separate column. The entries with "N/A" correspond to a situation in which the residue in that column was not ionized, and therefore $\Delta\Delta G_{\text{sol}}^{w\to p}$ values were not calculated. The reported pK_a^p values were used to calculate an approximate ratio of the conjugate base:acid forms using the Henderson–Hasselbalch equation with a pH value of 7.4, approximately the intracellular pH value [77]. The ratio $\frac{[A^-]}{[HA]}$ was solved for as:

$$\frac{[A^-]}{[HA]} = 10^{(pH - pK_a^p)} \quad (10)$$

As an example, Cys^0/His^+ (the third block in Table 1) refers to the cysteine being in the neutral, protonated state, with histidine in the protonated, positively charged state. In this permutation, it was possible to calculate pK_a^p values for Asp473, Asp350, and His471, though not Cys645.

The pK_a^p values can be used to estimate the population of a particular permutation or protonation microstate. Moving from the Cys^-/His^+ microstate (starting microstate in the proposed reverse protonation mechanism) to the Cys^0:His^0 microstate (starting microstate in the proposed substrate-assisted mechanism) can be accomplished in two hypothetical steps. First, a proton is removed from histidine (His^+) in the reaction

$$His^+ + Cys^- \rightarrow His^0 + Cys^- + H^+ \quad (11)$$

which gives an acid dissociation equilibrium (K_a) expression of

$$K_a = \frac{[Cys^- : His^0][H^+]}{[Cys^- : His^+]} = 10^{-pK_{a,His}^p} = 10^{-12.82} = 1.51 \times 10^{-13}. \quad (12)$$

The 12.82 in Eq. 12 corresponds to the pK_a^p value for histidine in the first block of Table 1. The second hypothetical step would be transferring a proton to cysteine (Cys^-) in the reaction

$$His^0 + Cys^- + H^+ \rightarrow His^0 + Cys^0. \quad (13)$$

The reverse of this reaction would correspond to the acid dissociation equilibrium, which would give

$$\frac{[Cys^- : His^0][H^+]}{[Cys^0 : His^0]} = 10^{-pK_{a,Cys}^p} = 10^{-11.80} = 1.58 \times 10^{-12}. \quad (14)$$

The 11.80 in Eq. 14 corresponds to the pK_a^p value for cysteine in the second block of Table 1. When the equilibrium expression in Eq. 12 is divided by the equilibrium expression in Eq. 14, the intermediate microstate (Cys^-:His^0) cancels and the expression becomes

$$\frac{[Cys^0 : His^0]}{[Cys^- : His^+]} = 10^{-1.02} = 0.095. \quad (15)$$

Thus, the calculations predict that the Cys^-: His^+ microstate predominates over the Cys^0: His^0 microstate by approximately a 10:1 ratio.

A similar comparison can be done for the Cys^0:His^+ microstate by the hypothetical addition of a proton to the Cys^0: His^0 microstate as shown in the reaction:

$$His^0 + Cys^0 + H^+ \rightarrow His^+ + Cys^0. \quad (16)$$

Again, the reverse of this reaction would correspond to the acid dissociation equilibrium, which would give an expression of

Table 1 $\Delta\Delta G_{sol}^{w\rightarrow p}$, pK_a^p and $\frac{[A^-]}{[HA]}$ ratios for the four different permutations of active site residues of PAD4 without BAG (1WD9)

Center residue	Asp473	Asp350	Cys645	His471
Permutation	Cys^-/His^+	Cys^-/His^+	Cys^-/His^+	Cys^-/His^+
$\Delta\Delta G_{sol}^{w\rightarrow p}$ of center residue (kcal/mol)	4.56 ± 0.70	5.13 ± 0.52	0.35 ± 0.82	−9.35 ± 0.71
pK_a^p of center residue	7.13 ± 0.51	7.54 ± 0.38	9.36 ± 0.60	*12.82* ± 0.52
$\frac{[A^-]}{[HA]}$	1.87	0.72	1.10×10^{-2}	3.79×10^{-6}
Permutation	Cys^-/His^0	Cys^-/His^0	Cys^-/His^0	Cys^-/His^0
$\Delta\Delta G_{sol}^{w\rightarrow p}$ of center residue (kcal/mol)	6.10 ± 0.59	10.25 ± 0.75	3.70 ± 0.49	N/A
pK_a^p of center residue	8.25 ± 0.43	11.28 ± 0.54	*11.80* ± 0.36	N/A
$\frac{[A^-]}{[HA]}$	0.14	1.33×10^{-4}	4.02×10^{-5}	N/A
Permutation	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+
$\Delta\Delta G_{sol}^{w\rightarrow p}$ of center residue (kcal/mol)	0.68 ± 0.85	−1.87 ± 0.40	N/A	−5.97 ± 0.49
pK_a^p of center residue	4.29 ± 0.62	2.44 ± 0.29	N/A	*10.35* ± 0.36
$\frac{[A^-]}{[HA]}$	1.28×10^3	9.20×10^4	N/A	1.11×10^{-3}
Permutation	Cys^0/His^0	Cys^0/His^0	Cys^0/His^0	Cys^0/His^0
$\Delta\Delta G_{sol}^{w\rightarrow p}$ of center residue (kcal/mol)	1.39 ± 1.53	7.39 ± 0.46	N/A	N/A
pK_a^p of center residue	4.81 ± 1.12	9.19 ± 0.33	N/A	N/A
$\frac{[A^-]}{[HA]}$	3.88×10^2	1.61×10^{-2}	N/A	N/A

All calculations had Asp473 and Asp350 negatively charged. The pK_a^p values in italics are those that were used in the microstate analyses. Standard deviations are shown for the solvation energy and pK_a^p calculations

$$K_a = \frac{[\mathrm{Cys}^0 : \mathrm{His}^0][\mathrm{H}^+]}{[\mathrm{Cys}^0 : \mathrm{His}^+]} = 10^{-\mathrm{pK}^p_{a,\mathrm{His}}} = 10^{-10.35} = 4.47 \times 10^{-11}. \quad (17)$$

The value of 10.35 corresponds to the pK^p_a value of histidine in block 3 of Table 1 (the $\mathrm{Cys}^0/\mathrm{His}^+$ permutation). At a cellular pH of 7.4, as suggested above, the expression for the ratio of microstates could be given as

$$10^{(\mathrm{pH}-\mathrm{pK}^p_a)} = \frac{[\mathrm{A}^-]}{[\mathrm{HA}]} = \frac{[\mathrm{Cys}^0 : \mathrm{His}^0]}{[\mathrm{Cys}^0 : \mathrm{His}^+]} = 10^{7.4-10.35} = 10^{-2.95} \approx 1.12 \times 10^{-3}. \quad (18)$$

As the predominance of the ion pair was only 10-fold more likely than the neutral system, the ratio found from Eq. 18 indicates the active site microstate with cysteine and histidine protonated is the most likely of the permutations in the absence of BAG, with a ratio of approximately 900:1 for Cys^0:His^+ versus Cys^0:His^0. This is not unreasonable, considering the presence of negative charges from the aspartate sidechains. It should be noted that in some permutations shown in Table 1, physically unreasonable pK^p_a values arise for the aspartic acids. This is due to a large concentration of negative charge in the active site, an extremely unlikely scenario, which greatly destabilizes the negative form of the aspartic acid, thus raising the pK^p_a values unrealistically, and contrary to experiments that indicate they are negatively charged [31, 39].

It should also be noted that in the ligand-free form of the enzyme, the predicted pK^p_a value for Cys645 of 9.36 when His471 is charged, predicted here to be the most populated microstate, agrees well (perhaps fortuitously) with a pK_a analysis by Li et al. of L-arginine deiminase (which also belongs to the GMSF family) which found the cysteine has a pK^p_a value of approximately 9.6 in the ligand-free form [21]. The authors of the study suggest that the cysteine pK^p_a value is shifted higher due to the proximity of the two catalytic aspartic acid residues (similar to the situation in PAD4).

An examination of the aspartic acids in the most populated microstate ($\mathrm{Cys}^0/\mathrm{His}^+$) shows predicted pK^p_a values of 4.29 (Asp473) and 2.44 (Asp350), consistent with both being primarily negatively charged at physiological pH, and values that are similar to their values in water. This study predicts that $\mathrm{Asp}^-/\mathrm{Asp}^-/\mathrm{Cys}^0/\mathrm{His}^+$ is the most populated microstate prior to substrate binding. The negative charges on the aspartic acid sidechains can coordinate with the positive charge of the entering guanidinium group on BAG through electrostatic attractions. Furthermore, with both aspartic acids creating a highly anionic active site, a point that the literature agrees with [3, 4], it is logical that a positively charged histidine would be favorable in the absence of BAG. This result is consistent with an experimental observation of DDAH [38].

It should be noted that the standard deviations for $\Delta\Delta G^{w\to p}_{\mathrm{sol}}$ were large in instances where $\Delta\Delta G^{w\to p}_{\mathrm{sol}}$ values were small, due to a fluctuation between small positive and negative numbers. As the values were all quite small when the standard deviations were large, the resultant pK^p_a values do not change dramatically from their solution pK^p_a values.

Table 2 presents results for PAD4 when it is bound to BAG, which was found to have invariably high pK^p_a values in all simulations. The pK^p_a values used in the calculations of microstate populations are italicized. A microstate analysis analogous to that done for Table 1 gives

$$\frac{[\mathrm{Cys}^0 : \mathrm{His}^0]}{[\mathrm{Cys}^- : \mathrm{His}^+]} = 10^{-5.66+9.38} = 10^{3.72} \approx 5,250 \quad (19)$$

Thus, the calculations predict that the Cys^0:His^0 microstate strongly predominates over the Cys^-:His^+ microstate in the presence of the positively charged substrate. Again, this should be compared to the likelihood of a Cys^0:His^+ microstate by the hypothetical addition of a proton to the Cys^0:His^0 microstate as shown in reaction 16. At a cellular pH of 7.4, as suggested above, the expression for the ratio of microstates could be given as

$$10^{(\mathrm{pH}-\mathrm{pK}^p_a)} = \frac{[\mathrm{A}^-]}{[\mathrm{HA}]} = \frac{[\mathrm{Cys}^0 : \mathrm{His}^0]}{[\mathrm{Cys}^0 : \mathrm{His}^+]} = 10^{7.4-2.59} = 10^{4.81} \approx 6.46 \times 10^4. \quad (20)$$

This population analysis indicates that Cys^0:His^0 is the most likely protonation microstate in PAD4 in the presence of BAG and is significantly more predominant than either Cys^0:His^+ or Cys^-:His^+.

A close examination of the predicted pK^p_a values for BAG reveals very large shifts from the solution pK_a value of approximately 12.5, with the largest shift observed in the $\mathrm{Cys}^-/\mathrm{His}^0/\mathrm{Bag}^+$ microstate. This state also has the two negatively charged aspartate sidechains, leading to a large concentration of negative charge, and perhaps a physically unrealistic scenario. This excess negative charge overstabilizes the positive charge on the BAG, leading to the large predicted pK^p_a value. As this state is probably not achievable, it leads to an unrealistic estimate of the pK^p_a value.

Plotting the fraction of enzyme that has a deprotonated histidine and a protonated cysteine in the presence of the positively charged substrate using the calculated pK^p_a values of 2.59 for His471 and 9.38 for Cys645 reveals that the vast majority of enzyme is in the active form over a large pH range, including cellular pH. This is shown in Fig. 6, where the yellow area under the curves for His471 and Cys645 is the fraction of enzyme in the active form (having

Table 2 $\Delta\Delta G^{w\to p}$, pK_a^p and $\frac{[A^-]}{[HA]}$ ratios for the active site residues of the PAD4 enzyme containing the substrate benzoyl-L-arginine amide (BAG) under four different permutations

Center Residue	Asp 471	Asp 350	Bag	Cys645	His471
Permutation	$Cys^-/His^+/Bag^+$	$Cys^-/His^+/Bag^+$	$Cys^-/His^+/Bag^+$	$Cys^-/His^+/Bag^+$	$Cys^-/His^+/Bag^+$
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	2.71 ± 0.71	−8.18 ± 1.02	−11.03 ± 1.43	−3.85 ± 0.47	0.46 ± 0.93
pK_a^p of center residue	5.78 ± 0.52	−2.16 ± 0.75	20.54 ± 1.04	6.29 ± 0.34	*5.66* ± 0.68
$\frac{[A^-]}{[HA]}$	42.07	3.66×10^{9}	7.22×10^{-14}	12.78	54.67
Permutation	$Cys^-/His^0/Bag^+$	$Cys^-/His^0/Bag^+$	$Cys^-/His^0/Bag^+$	$Cys^-/His^0/Bag^+$	$Cys^-/His^0/Bag^+$
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	5.84 ± 1.16	−0.36 ± 0.80	−23.36 ± 0.72	0.38 ± 0.72	N/A
pK_a^p of center residue	8.06 ± 0.84	3.54 ± 0.58	29.54 ± 0.52	*9.38* ± 0.52	N/A
$\frac{[A^-]}{[HA]}$	0.22	7.31×10^{3}	7.25×10^{-23}	1.05×10^{-2}	N/A
Permutation	$Cys^0/His^+/Bag^+$	$Cys^0/His^+/Bag^+$	$Cys^0/His^+/Bag^+$	$Cys^0/His^+/Bag^+$	$Cys^0/His^+/Bag^+$
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	−5.28 ± 0.80	−14.56 ± 0.76	−3.69 ± 1.19	N/A	4.68 ± 0.45
pK_a^p of center residue	−0.05 ± 0.58	−6.82 ± 0.55	15.19 ± 0.87	N/A	*2.59* ± 0.33
$\frac{[A^-]}{[HA]}$	2.84×10^{7}	1.67×10^{14}	1.62×10^{-8}	N/A	6.52×10^{4}
Permutation	$Cys^0/His^0/Bag^+$	$Cys^0/His^0/Bag^+$	$Cys^0/His^0/Bag^+$	$Cys^0/His^0/Bag^+$	$Cys^0/His^0/Bag^+$
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	−3.28 ± 0.82	−5.86 ± 1.06	−14.40 ± 1.41	N/A	N/A
pK_a^p of center residue	1.41 ± 0.60	−0.47 ± 0.77	23.00 ± 1.03	N/A	N/A
$\frac{[A^-]}{[HA]}$	9.75×10^{5}	7.49×10^{7}	2.50×10^{-16}	N/A	N/A

Again, Asp471 and Asp350 are both negatively charged in all calculations. The pK_a^p values in italics are those that were used in the microstate analyses. Standard deviations are shown for the solvation energy and pK_a^p calculations

the sidechains of both His471 and Cys645 in their neutral form). The large range where the enzyme exists in the optimum form is consistent with observations of a relatively high k_{cat} between pH values of approximately 7–9 [37]. These data provide a different view compared to previously published results of iodoacetamide inactivation and solvent kinetic isotope effect studies [37]. However, it is difficult to compare the results of this work with the iodoacetamide inactivation, as iodoacetamide does not contain the positively charged guanidinium group that the natural substrate possesses, and which is shown here to strongly affect the pK_a^p value of His471. Similarly, interpretation of solvent kinetic isotope effects depends on a number of different parameters, particularly in the presence of thiols [78, 79], and a discussion of this is outside the scope of this work.

It is important here to address changes between the two crystal structures to assess the effect the substrate has on the active site residues. This comparison is shown in Table 3. The permutation used for comparison is the Cys^0/His^+ because this is the condition that the active site is proposed to be in for the unbound enzyme. However, comparing other permutations between the two crystal structures yields similar results.

Upon the addition of the substrate, the pK_a^p of histidine drops significantly and the $\frac{[A^-]}{[HA]}$ ratio increases from 0.0011 to 6.52×10^4. This suggests the histidine changes from being mostly protonated in the absence of the substrate, to being mostly deprotonated upon substrate binding. This is a logical finding; the ionized histidine would be destabilized by the presence of the positively charged guanidinium group of the substrate in close proximity. Since the substrate has an invariably high pK_a^p, much higher than that of histidine, BAG will remain positively charged, and histidine will be deprotonated and neutral. This observation supports the neutral system (corresponding to the starting state of the substrate-assisted mechanism). Snapshots taken from simulations of the active site with substrate present, as in Fig. 7, show that in the Cys^0:His^0 microstate the atomic distances from the substrate to the aspartate sidechains are well suited for hydrogen bonding (Fig. 7).

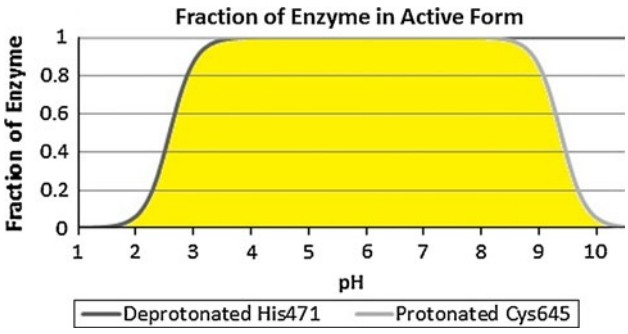

Fig. 6 The fraction of active enzyme is shown in *yellow*. This corresponds to having deprotonated His471 and protonated Cys645, the starting protonation state in the substrate-assisted mechanism

Table 3 $\Delta\Delta G^{w\to p}$, pK_a^p and $\frac{[A^-]}{[HA]}$ ratios comparing the bound and unbound crystal structures of PAD4

Center	Asp 471	Asp 350	Bag	Cys	His
Permutation	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+
Substrate	Absent	Absent	Absent	Absent	Absent
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	0.68	−1.87	N/A	N/A	−5.97
pK_a^p	4.29	2.44	N/A	N/A	10.35
$\frac{[A^-]}{[HA]}$	1.28×10^3	9.20×10^4	N/A	N/A	1.11×10^{-3}
Permutation	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+	Cys^0/His^+
Substrate	Bag^+	Bag^+	Bag^+	Bag^+	Bag^+
$\Delta\Delta G_{sol}^{w\to p}$ of center residue (kcal/mol)	−5.28	−14.56	−3.69	N/A	4.68
pK_a^p	−0.05	−6.82	15.19	N/A	2.59
$\frac{[A^-]}{[HA]}$	2.84×10^7	1.67×10^{14}	1.62×10^{-8}	N/A	6.52×10^4

Another finding is that the pK_a^p values of the aspartic acid residues dropped, indicating they remain deprotonated and negatively charged upon addition of the substrate, suggesting the negative charges facilitate coordination with the positive substrate. This allows the substrate to be held in place during the reaction and is an expected result. The role of the aspartic acids is not under controversy in the literature [31, 34, 37, 39].

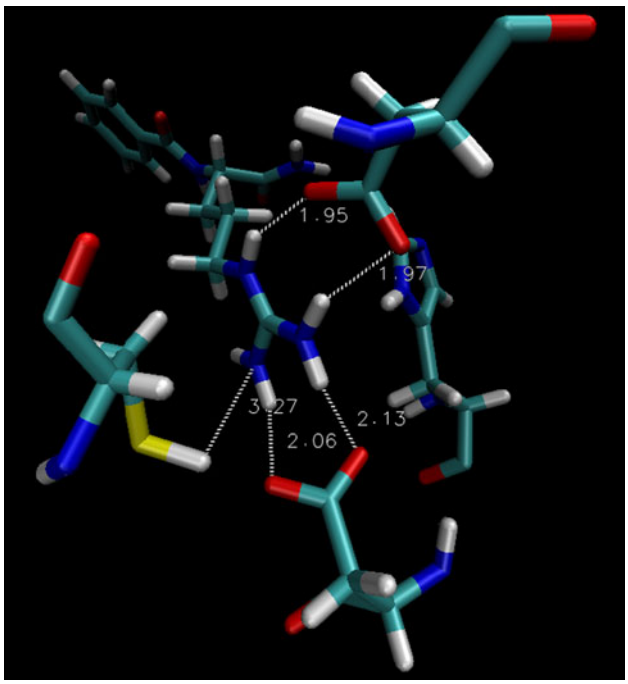

Fig. 7 Snapshot of the active site of PAD4 in the neutral state with BAG bound, showing distances of hydrogen atoms to heavy atoms in Angstroms. The *colors* of the atoms are as follows: *red* for oxygen, *green* for carbon, *blue* for nitrogen, *white* for hydrogen, and *yellow* for sulfur

4 Summary and conclusions

Using the software package MOLARIS, the change in Gibbs free energy as each active site residue of PAD4 is moved from water to protein was calculated, and these values were used to determine pK_a^p values for each residue. From these pK_a^p values, the microstate populations could be estimated, giving insight into the most stable permutations of the active site residues. Results show that in the absence of substrate, both cysteine and histidine are protonated, indicating that cysteine and histidine are neutral and positively charged, respectively. Both aspartic acid residues appear to be deprotonated and negatively charged.

In the presence of substrate, results show that cysteine and histidine are protonated and deprotonated, respectively, and are therefore both neutral. The substrate itself has invariably high pK_a^p values, indicating it is always protonated. Both aspartic acid residues appear deprotonated and negatively charged to facilitate binding to the substrate. A likely explanation for the condition of active site with bound substrate is that the close proximity of the positively charged substrate causes histidine to prefer the neutral state, and therefore could not stabilize an ion pair with cysteine. Furthermore, the cysteine and histidine, being 6 Å apart, are probably too far from each other to be stabilized as an ion pair.

This study provides support for a substrate-assisted mechanism which begins with histidine and cysteine in the neutral state when a positively charged substrate is bound. Knowing the mechanism of this enzyme is crucial for development of small-molecule inhibitors because it allows us to understand important properties of the active site such as polarity and hydrogen bond donors and acceptors. This information could lead to the development of inhibitors that would most efficiently bind. Much work is still necessary to fully understand the nature and mechanism of

PAD4; however, its role in crippling diseases such as RA makes further work in this area very relevant.

Acknowledgments This work was supported by the University of Colorado Colorado Springs. Dr. Arieh Warshel and Dr. Zhen Tao Chu at the University of Southern California are thanked for access to the MOLARIS software and for helpful discussions. Dr. James Vivian is also thanked for helpful discussions.

References

1. Hagiwara T, Nakashima K, Hirano H, Senshu T, Yamada M (2002) Biochem Biophys Res Commun 290:979–983
2. Nakashima K, Hagiwara T, Yamada M (2002) J Biol Chem 277:49562–49568
3. Hagiwara T, Hidaka Y, Yamada M (2005) Biochemistry 44:5827–5834
4. Wang Y, Wysocka J, Sayegh J, Lee Y-H, Perlin JR, Leonelli L, Sonbuchner LS, McDonald CH, Cook RG, Dou Y, Roeder RG, Clarke S, Stallcup MR, Allis CD, Coonrod SA (2004) Science, vol 306. Washington, DC, pp 279–283
5. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, Schneider R, Gregory PD, Tempst P, Bannister AJ, Kouzarides T (2004) Cell, 118. MA, Cambridge, pp 545–553
6. Wang Y, Li M, Stadler S, Correll S, Li P, Wang D, Hayama R, Leonelli L, Han H, Grigoryev SA, Allis CD, Coonrod SA (2009) J Cell Biol 184:205–213
7. Guo Q, Fast W (2011) J Biol Chem 286:17069–17078
8. Li P, Yao H, Zhang Z, Li M, Luo Y, Thompson PR, Gilmour DS, Wang Y (2008) Mol Cell Biol 28:4745–4758
9. Jones JE, Causey CP, Knuckley B, Slack-Noyes JL, Thompson PR (2009) Curr Opin Drug Discov Devel 12:616–627
10. Schellekens GA, De Jong BAW, Van Den Hoogen FHJ, Van De Putte LBA, Van Venrooij WJ (1998) J Clin Investig 101:273–281
11. Anzilotti C, Pratesi F, Tommasi C, Migliorini P (2010) Autoimmun Rev 9:158–160
12. Vossenaar ER, Zendman AJW, van Venrooij WJ (2004) Arthr Res Ther 6:1–5
13. Shirai H, Blundell TL, Mizuguchi K (2001) Trends Biochem Sci 26:465–468
14. Lu X, Galkin A, Herzberg O, Dunaway-Mariano D (2004) J Am Chem Soc 126:5374–5375
15. Galkin A, Kulakova L, Sarikaya E, Lim K, Howard A, Herzberg O (2004) J Biol Chem 279:14001–14008
16. Das K, Butler GH, Kwiatkowski V, Clark AD, Yadav P, Arnold E (2004) Structure, vol 12. MA, Cambridge, pp 657–667
17. Galkin A, Lu X, Dunaway-Mariano D, Herzberg O (2005) J Biol Chem 280:34080–34087
18. Lu X, Li L, Feng X, Wu Y, Dunaway-Mariano D, Engen JR, Mariano PS (2005) J Am Chem Soc 127:16412–16413
19. Lu X, Li L, Wu R, Feng X, Li Z, Yang H, Wang C, Guo H, Galkin A, Herzberg O, Mariano PS, Martin BM, Dunaway-Mariano D (2006) Biochemistry 45:1162–1172
20. Li L, Li Z, Chen D, Lu X, Feng X, Wright EC, Solberg NO, Dunaway-Mariano D, Mariano PS, Galkin A, Kulakova L, Herzberg O, Green-Church KB, Zhang L (2008) J Am Chem Soc 130:1918–1931
21. Li L, Li Z, Wang C, Xu D, Mariano PS, Guo H, Dunaway-Mariano D (2008) Biochemistry 47:4721–4732
22. Ke Z, Guo H, Xie D, Wang S, Zhang Y (2011) J Phys Chem B 115:3725–3733
23. Li J, Xu P, Jiao Q (2009) Gongye Weishengwu 39:1–5
24. Li Z, Kulakova L, Li L, Galkin A, Zhao Z, Nash TE, Mariano PS, Herzberg O, Dunaway-Mariano D (2009) Bioorg Chem 37:149–161
25. Murray-Rust J, Leiper J, McAlister M, Phelan J, Tilley S, Santa Maria J, Vallance P, McDonald N (2001) Nat Struct Biol 8:679–683
26. Humm A, Fritsche E, Steinbacher S, Huber R (1997) EMBO J 16:3373–3385
27. Fritsche E, Bergner A, Humm A, Piepersberg W, Huber R (1998) Biochemistry 37:17664–17672
28. Linsky T, Fast W (2010) Biochim Biophys Acta Proteins Proteom 1804:1943–1953
29. Muenchhoff J, Siddiqui KS, Poljak A, Raftery MJ, Barrow KD, Neilan BA (2010) FEBS J 277:3844–3860
30. Shirai H, Mokrab Y, Mizuguchi K (2006) Proteins: Struct, Funct, Bioinf 64:1010–1023
31. Arita K, Hashimoto H, Shimizu T, Nakashima K, Yamada M, Sato M (2004) Nat Struct Mol Biol 11:777–783
32. Vossenaar ER, Zendman AJW, van Venrooij WJ, Pruijn GJM (2003) BioEssays 25:1106–1118
33. Kearney PL, Bhatia M, Jones NG, Yuan L, Glascock MC, Catchings KL, Yamada M, Thompson PR (2005) Biochemistry 44:10570–10582
34. Ke Z, Zhou Y, Hu P, Wang S, Xie D, Zhang Y (2009) J Phys Chem B 113:12750–12758
35. Arita K, Shimizu T, Hashimoto H, Hidaka Y, Yamada M, Sato M (2006) Proc Natl Acad Sci USA 103:5291–5296
36. Ke Z, Wang S, Xie D, Zhang Y (2009) J Phys Chem B 113:16705–16710
37. Knuckley B, Bhatia M, Thompson PR (2007) Biochemistry 46:6578–6587
38. Stone EM, Costello AL, Tierney DL, Fast W (2006) Biochemistry 45:5618–5630
39. Leopoldini M, Marino T, Toscano M (2008) Theoret Chem Acc 120:459–466
40. Jimenez-Morales D, Liang J, Eisenberg B (2012) Eur Biophys J EBJ 41:449–460
41. Ma S, Devi-Kesavan LS, Gao J (2007) J Am Chem Soc 129:13633–13645
42. Mladenovic M, Fink RF, Thiel W, Schirmeister T, Engels B (2008) J Am Chem Soc 130:8696–8705
43. Rastogi VK, Girvin ME (1999) Nature 402:263–268
44. Isom DG, Castaneda CA, Cannon BR, Garcia-Moreno B (2011) Proc Natl Acad Sci USA 108:5260–5265
45. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, Garcia-Moreno EB (2000) Biophys J 79:1610–1620
46. Harms MJ, Castaneda CA, Schlessman JL, Sue GR, Isom DG, Cannon BR, Garcia-Moreno EB (2009) J Mol Biol 389:34–47
47. Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, Bertrand G-ME (2007) Biophys J 92:2041–2053
48. Fitch CA, Karp DA, Lee KK, Stites WE, Lattman EE, Garcia-Moreno EB (2002) Biophys J 82:3289–3304
49. Garcia-Moreno EB, Dwyer JJ, Gittis AG, Lattman EE, Spencer DS, Stites WE (1997) Biophys Chem 64:211–224
50. Stites WE, Gittis AG, Lattman EE, Shortle D (1991) J Mol Biol 221:7–14
51. Warshel A, Dryga A (2011) Proteins: Struct, Funct, Bioinf 79:3469–3484
52. Schutz CN, Warshel A (2001) Proteins Struct Funct Genet 44:400–417
53. King G, Lee FS, Warshel A (1991) J Chem Phys 95:4366–4377
54. Whitten ST, Garcia-Moreno EB, Hilser VJ (2005) Proc Natl Acad Sci USA 102:4282–4287
55. Crystal structure of human peptidylarginine deiminase type4 (PAD4) in complex with benzoyl-L-arginine amide.

http://www.pdb.org/pdb/explore/explore.do?structureId=1wda. Last accessed 10 Jan 2012
56. Calcium bound form of human peptidylarginine deiminase type4 (PAD4). http://www.pdb.org/pdb/explore/explore.do?structureId=1wd9. Last accessed 10 Jan 2012
57. SPARTAN '08 Wavefunction, Inc. 18401 Von Karman, Suite 370, Irvine, CA 92612. http://www.wavefun.com. Last accessed 21 July 2012
58. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA Jr, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, Van VT, 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, Woodcock HL III, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer HF III, Kong J, Krylov AI, Gill PMW, Head-Gordon M (2006) Phys Chem Chem Phys 8:3172–3191
59. MOLARIS version beta 9.05 http://futura.usc.edu. Last accessed 21 July 2012
60. Lee FS, Chu ZT, Warshel A (1993) J Comput Chem 14:161–185
61. Sham YY, Chu ZT, Warshel A (1997) J Phys Chem B 101:4458–4472
62. Warshel A, Naray-Szabo G, Sussman F, Hwang JK (1989) Biochemistry 28:3629–3637
63. Kato M, Pisliakov AV, Warshel A (2006) Proteins: Struct, Funct, Bioinf 64:829–844
64. Warshel A, Sharma PK, Kato M, Parson WW (2006) Biochim Biophys Acta Proteins Proteom 1764:1647–1676
65. Rosta E, Klaehn M, Warshel A (2006) J Phys Chem B 110:2934–2941
66. Kato M, Warshel A (2006) J Phys Chem B 110:11566–11570
67. Langen R, Brayer GD, Berghuis AM, McLendon G, Sherman F, Warshel A (1992) J Mol Biol 224:589–600
68. Xiang Y, Oelschlaeger P, Florian J, Goodman MF, Warshel A (2006) Biochemistry 45:7036–7048
69. Warshel A (1981) Biochemistry 20:3167–3177
70. Chu, ZT (2009) MOLARIS: Version 9.11 theoretical background and Practical Examples http://futura.usc.edu/programs/doc/theory_molaris_9.11.pdf. Last accessed 21 July 2012
71. Burykin A, Schutz CN, Villa J, Warshel A (2002) Proteins Struct Funct Genet 47:265–280
72. Sham YY, Muegge I, Warshel A (1998) Biophys J 74:1744–1753
73. Lim C, Bashford D, Karplus M (1991) J Phys Chem 95:5610–5620
74. Jorgensen WL, Briggs JM (1989) J Am Chem Soc 111:4190–4197
75. Warshel A (1979) J Phys Chem 83:1640–1652
76. Warshel A, Russell ST, Churg AK (1984) Proc Natl Acad Sci USA 81:4785–4789
77. Xiong Y, Lu H-T, Zhan C-G (2008) J Comput Chem 29:1259–1267
78. Schneck JL, Villa JP, McDevitt P, McQueney MS, Thrall SH, Meek TD (2008) Biochemistry 47:8697–8710
79. Karsten WE, Lai C-J, Cook PF (1995) J Am Chem Soc 117:5914–5918