

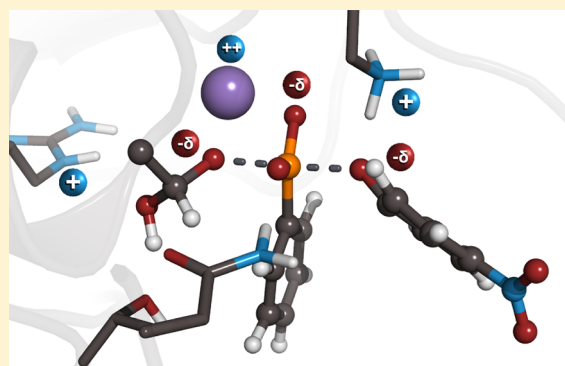
# Cooperative Electrostatic Interactions Drive Functional Evolution in the Alkaline Phosphatase Superfamily

Alexandre Barrozo,<sup>‡</sup> Fernanda Duarte,<sup>‡</sup> Paul Bauer, Alexandra T. P. Carvalho, and Shina C. L. Kamerlin\*

Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, SE-751 24, Uppsala, Sweden

## Supporting Information

**ABSTRACT:** It is becoming widely accepted that catalytic promiscuity, i.e., the ability of a single enzyme to catalyze the turnover of multiple, chemically distinct substrates, plays a key role in the evolution of new enzyme functions. In this context, the members of the alkaline phosphatase superfamily have been extensively studied as model systems in order to understand the phenomenon of enzyme multifunctionality. In the present work, we model the selectivity of two multiply promiscuous members of this superfamily, namely the phosphonate monoester hydrolases from *Burkholderia caryophylli* and *Rhizobium leguminosarum*. We have performed extensive simulations of the enzymatic reaction of both wild-type enzymes and several experimentally characterized mutants. Our computational models are in agreement with key experimental observables, such as the observed activities of the wild-type enzymes, qualitative interpretations of experimental pH-rate profiles, and activity trends among several active site mutants. In all cases the substrates of interest bind to the enzyme in similar conformations, with largely unperturbed transition states from their corresponding analogues in aqueous solution. Examination of transition-state geometries and the contribution of individual residues to the calculated activation barriers suggest that the broad promiscuity of these enzymes arises from cooperative electrostatic interactions in the active site, allowing each enzyme to adapt to the electrostatic needs of different substrates. By comparing the structural and electrostatic features of several alkaline phosphatases, we suggest that this phenomenon is a generalized feature driving selectivity and promiscuity within this superfamily and can be in turn used for artificial enzyme design.



## INTRODUCTION

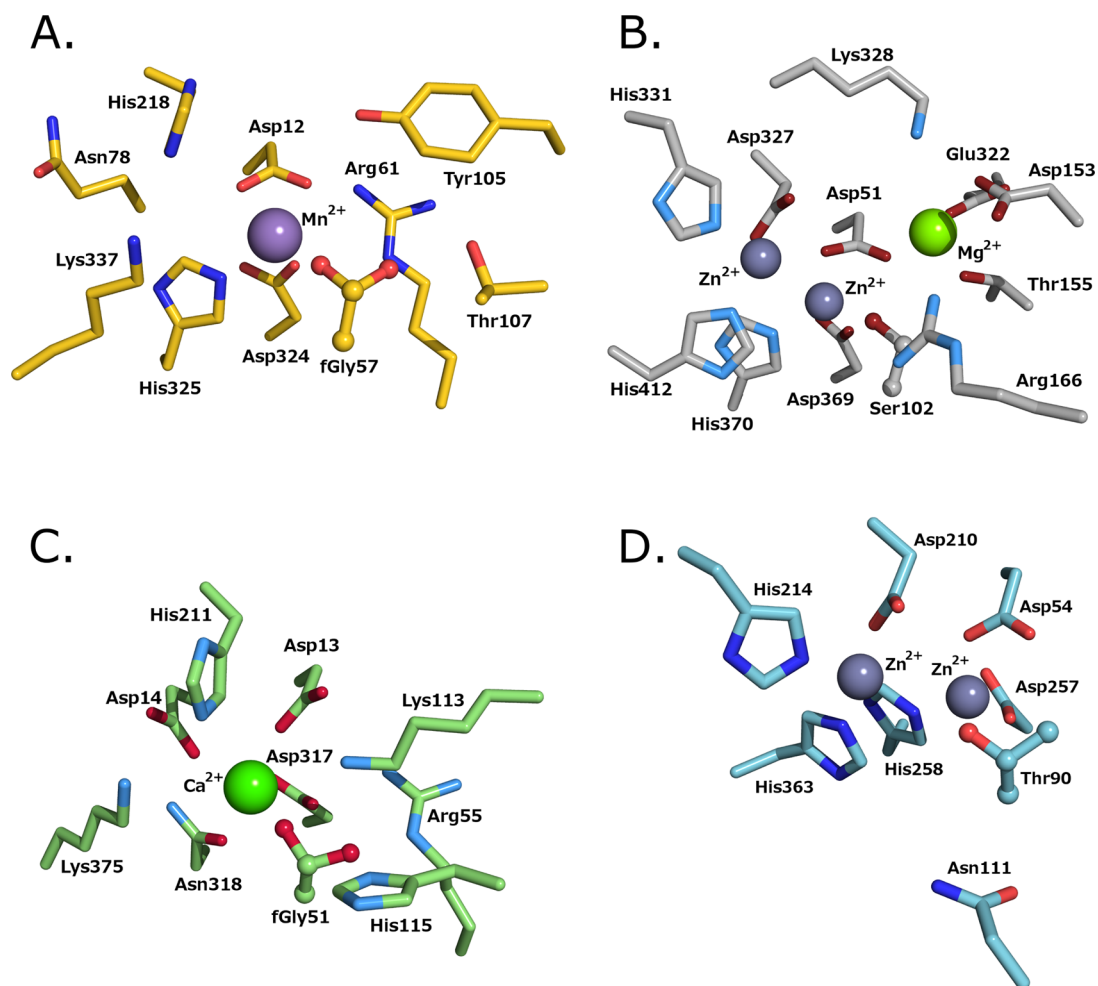
In recent years,<sup>1</sup> it has become widely accepted that catalytic promiscuity, i.e., the ability of many enzymes to catalyze the turnover of multiple chemically distinct substrates, plays a key role in the evolution of new functions, allowing for rapid responses to environmental changes.<sup>2,3</sup> Furthermore, interest in this phenomenon has exploded as it has been increasingly shown to be a powerful tool for gaining knowledge not just into the process of natural functional evolution,<sup>2</sup> but also as a factor that can be exploited in effective artificial enzyme design.<sup>1,3,4</sup> Such promiscuity appears to be highly pronounced among many phosphotransferases, such as the recently evolved bacterial phosphotriesterase (PTE),<sup>5</sup> serum paraoxonase 1 (PON1),<sup>6</sup> and members of the alkaline phosphatase (AP) superfamily,<sup>7–9</sup> to name a few examples. This latter superfamily has additionally played a central role as a model system for understanding enzyme catalytic promiscuity,<sup>7,9–13</sup> i.e., the ability of a given enzyme to catalyze more than one distinct chemical reaction.

The characterized members of the AP superfamily are highly promiscuous hydrolytic enzymes capable of interchangeable cleavage of P–O, S–O, and P–C bonds.<sup>14,15</sup> That is, they have been shown to catalyze the hydrolysis of a range of substrates

that differ in the nature of their TS solvation and protonation patterns, and thus in their requirements for efficient catalysis (see discussion in refs 16–18). Furthermore, all known AP superfamily members are metallohydrolases that employ similar catalytic scaffolds, which are comprised of at least one divalent metal ion in their respective active sites (Figure 1). This metal ion plays an important role in activating the nucleophile, which is generally thought to be an alcohol or alkoxide depending on the particular superfamily member,<sup>15</sup> by increasing the concentration of its active deprotonated form. Additionally, while there are a number of similarities between different known members of the superfamily, there are also broad differences in their metal requirements, overall structures, and specific choice of nucleophile, which can in turn be linked to changes in specificity patterns.<sup>11</sup> Despite these differences, a particular hallmark of this superfamily is crosswise-promiscuity, in that the native substrate for one member of the superfamily is often a promiscuous substrate for another,<sup>11,15</sup> in some cases with high (and almost comparable) proficiencies toward both the native and promiscuous substrates.<sup>8,9,12</sup> As a result, these

Received: April 16, 2015

Published: June 19, 2015



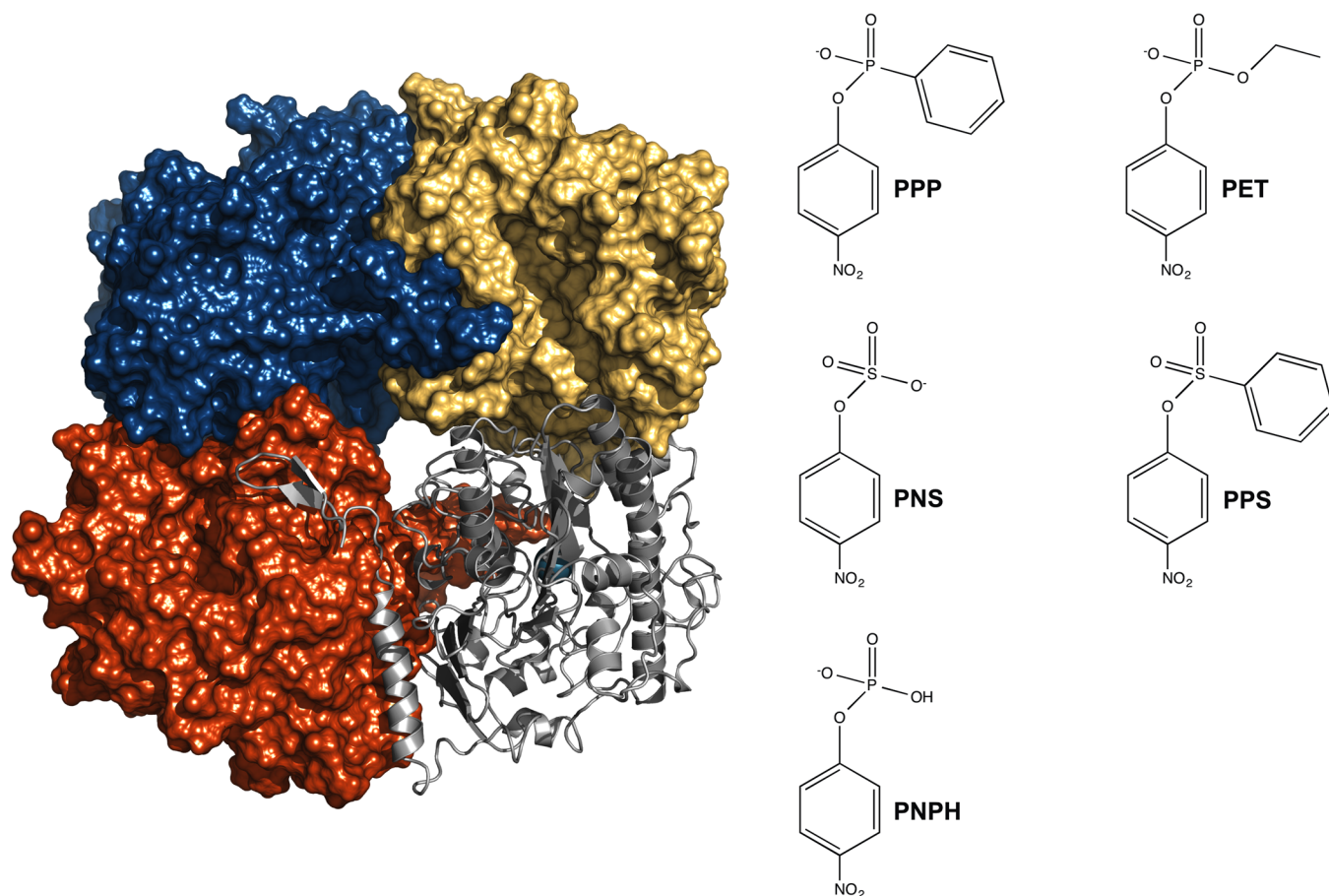
**Figure 1.** An active site comparison of selected members of the AP superfamily. Shown here are the active sites of (A) *Rhizobium leguminosarum* phosphonate monoesterase (PDB code 2VQR),<sup>8</sup> (B) *Escherichia coli* alkaline phosphatase (1ALK),<sup>20</sup> (C) *Pseudomonas aeruginosa* arylsulfatase (1HDH)<sup>19</sup> and (D) *Xanthomonas axonopodis* nucleotide pyrophosphatase/phosphodiesterase (2GSN).<sup>21</sup> The figure highlights the presence of divalent metal ions as well as the conservation of some of the residues surrounding them.

enzymes provide a perfect showcase to generate a systematic roadmap of the process of functional evolution within an enzyme superfamily, as well as a broader model system for understanding the evolution of phosphohydrolase activity.<sup>15</sup>

Among the different superfamily members, the name giving member AP<sup>20,22–25</sup> as well as the very closely related nucleotide pyrophosphatase/phosphodiesterase (NPP)<sup>21,25–27</sup> have been the subject of extensive scrutiny. A lesser-studied subset of enzymes that stand out in this superfamily are those classified as phosphonate monoester hydrolases (PMHs), such as the enzymes from *Rhizobium leguminosarum* (RIPMH)<sup>8</sup> and *Burkholderia caryophylli* (BcPMH).<sup>12</sup> These highly promiscuous enzymes efficiently promote the hydrolysis of at least five different substrate classes (Figure 2) and stand out in particular as their promiscuous phosphodiesterase activity is almost as efficient as their native phosphonate monoesterase activity<sup>8,12</sup> (Table S1); note that the PTE activity reported in this work is ambiguous, as discussed in the Results and Discussion. Moreover, these PMHs provide the first example of biological PMH activity and are the only currently known enzyme capable of catalyzing the hydrolysis of xenobiotic sulfonate esters by direct S–O cleavage.<sup>11,12</sup> Note also that both enzymes are large homo tetramers with ~56 kDa subunits and have extremely large binding pockets (~10 × 20 Å<sup>2</sup> wide and 15 Å deep).<sup>8,12</sup>

Therefore, one would assume that such enzymes could easily accommodate a range of substrates of different shapes and sizes. Perhaps unsurprisingly, therefore, both PMHs are moderately efficient catalysts for the hydrolysis of the compounds shown in Figure 2 ( $k_{\text{cat}}/K_{\text{M}}$  values in the range of 10<sup>3</sup>–10<sup>4</sup>, see Table S1) and, in the case of RIPMH, apparently only marginally affected by mutations of the key active site residues with presumably multiple catalytic backups present in the active site (Table S2) that can take over the role of the mutated residues.

A closely evolutionarily related enzyme in the AP superfamily is the arylsulfatase from *Pseudomonas aeruginosa* (PAS) (Figure 1C).<sup>19</sup> This enzyme only shares about 27% sequence similarity to RIPMH but has high structural similarity, in that 64% of the residues between the two enzymes structurally align with an RMSD of 2.54 Å.<sup>8</sup> This enzyme also has recently been the subject of extensive experimental<sup>9,28</sup> and computational<sup>13,29</sup> studies. Both PAS and the PMHs contain a mononuclear metal center with distorted octahedral conformation, which is most likely Mn<sup>2+</sup> in the PMHs<sup>8,12</sup> and Ca<sup>2+</sup> in PAS.<sup>19</sup> In addition, all three enzymes use an unusual geminal diol nucleophile<sup>8,12,28</sup> (Figure 1), a feature they share with all known sulfatases.<sup>30</sup> This noncanonical residue is a post-translationally modified cysteine or serine, which is first converted to an aldehyde and then hydrated to give rise to its active form.<sup>31</sup> Despite these apparent



**Figure 2.** Structure of *RIPMH* (PDB ID: 2VQR) and the corresponding substrates studied in this work.<sup>8</sup> Both *RIPMH* and *BcPMH* are dimers of dimers, in which the monomeric units of each dimer communicate with its corresponding oligomeric unit through the C-terminal loop highlighted in this figure (which is in turn an adaptation from an analogous figure presented in ref 12). This loop reaches into the adjacent active site, helping position key catalytic residues.<sup>8</sup> The substrates studied in this work are phenyl *p*-nitrophenyl phosphonate (PPP), ethyl *p*-nitrophenyl phosphate (PET), *p*-nitrophenyl sulfate (PNS), phenyl *p*-nitrophenyl sulfonate (PPS), and the *p*-nitrophenyl phosphate monoanion (PNPH).

similarities, the two PMHs and PAS have very different specificity patterns. That is, PMHs are phosphonate monoesterases, while PAS is primarily a sulfatase, although all three enzymes have relatively low discrimination between native and promiscuous substrates.<sup>8,9,12,28</sup> In contrast, other superfamily members such as AP and NPP have dinuclear zinc centers in their catalytic sites (AP also possesses a third metal ion that appears to play an important role in determining the activity)<sup>23</sup> and utilize ionized serine or threonine residues as nucleophiles, respectively<sup>10</sup> (Figure 1). Thus, a direct atomic-level comparison between individual AP superfamily members and also related promiscuous phosphatases can provide better and broader understanding of the features that drive selectivity and promiscuity in these highly multifunctional systems.

In the present work, we have performed an extensive number of empirical valence bond (EVB)<sup>32,33</sup> simulations (total simulation time of  $\sim 4 \mu\text{s}$ ) of both the native and several characterized promiscuous activities of *BcPMH* and *RIPMH*, reproducing key experimental observables such as activation barriers, qualitative predictions from the pH-rate profiles for *BcPMH* activity, and the effect of mutations on *RIPMH* activity.<sup>8,12</sup> We demonstrate that, despite their broad promiscuity, the PMHs studied in this work hydrolyze all substrates through a unified mechanism with similar substrate binding positions, transition states, and electrostatic contributions to transition-state stabilization. Additionally, we showcase

the importance of compensatory and cooperative electrostatic effects, which allow for an electrostatically flexible active site environment that can accommodate a range of substrates with different charge distributions, transition-state geometries, and requirements for efficient catalysis.

Finally, in order to test whether these observations are general to other members of the superfamily, we provide a detailed comparison of a range of AP superfamily members, in terms of active site shape, volume, and polarity. From this analysis we find a strong correlation between these properties and both substrate charge preference and number of known promiscuous activities, once again emphasizing the central role of the electrostatic environment of the active site in determining enzyme specificity and promiscuity. It is commonly accepted that enzymes achieve their tremendous catalytic proficiencies through an exquisite network of interactions that preferentially stabilize their transition states over their ground states,<sup>34</sup> and it has been argued that this is achieved through preorganization of the catalytic residues into an optimal conformation for transition-state stabilization.<sup>35,36</sup> This has been demonstrated for a wide range of systems through both experimental and computational work.<sup>37–39</sup> We illustrate here that while having an optimal electrostatic environment is clearly important to the catalysis of these enzymes toward individual substrates, one should also take into account the cooperativity between these residues, where the

effect of the combined electrostatic environment from all relevant residues on the transition state stabilization is greater than the effect of each residue determined individually. We demonstrate that this cooperativity renders the active site electrostatic environment sufficiently flexible to accommodate a broader range of substrates with different electrostatic needs for efficient catalysis (without necessarily altering either substrate binding position or enzyme conformation). Additionally, our comparative analysis of different alkaline phosphatases shows that the higher the number of polar active site residues, the greater the propensity toward catalytic promiscuity. This highlights the importance of such cooperative electrostatic interactions as a common feature to functional evolution among members of the AP superfamily,<sup>13</sup> illustrating the power of subtle amino acid substitutions to drive very different solutions for the same chemical problem.

## METHODOLOGY

**Initial System Setup.** Initial structures for both *RIPMH*<sup>8</sup> and *BcPMH*<sup>12</sup> (1.42 and 2.40 Å resolution, respectively) were obtained from the Protein Data Bank<sup>40</sup> (accession codes 2VQR and 2W8S). As the deposited structure for *RIPMH* contains only the monomeric unit without the transformation matrix, the structure of the full tetramer was obtained directly from the authors.<sup>8</sup> Potential flips of histidine, asparagine, and glutamine side chains were evaluated using the MolProbity server,<sup>41</sup> and those suggested by the software were applied to the structure. In all cases, the substrates were placed manually in the active site in such a way to optimize nonbonded interactions between the substrate and nearby amino acid side chains, including charge–charge, hydrogen-bonding and hydrophobic interactions. Structures for the corresponding Q13A, N78A, Y105A, T107A, H218A, and K337A variants presented in ref 8 were generated by manual truncation of the relevant side chains starting from the wild-type crystal structure, and structures were equilibrated using the same protocol as for the wild-type enzymes in order to allow the active site to adapt to the perturbation introduced.

Both PMHs are metalloenzymes with a single metal per active site of the tetramer (4 total), and the most likely candidate for this role has been identified as being a divalent manganese ion.<sup>8,12</sup> We recently presented a set of force-field-independent parameters to describe a range of alkali earth and transition-metal centers<sup>42</sup> based on Åqvist and Warshel's original cationic dummy model,<sup>43</sup> which describes the metal as a delocalized charge spread over a number of dummy atoms placed around the metal center (in this case six particles in octahedral coordination, as shown in Figure S1). These particles are bonded to the central atom and to each other, and the frame is allowed to freely rotate in its coordination sphere without the need for external constraints or artificial bonds. We demonstrated<sup>42</sup> that this model also allows one to capture subtle structural effects upon metal substitution without the need for the artificial restraints that need to be imposed in a fully bonded model, while simultaneously capturing key electrostatic properties of the metal center. We have successfully used our Ca<sup>2+</sup> model in simulations of the selectivity of PON1,<sup>44</sup> and the Mn<sup>2+</sup> model presented in our original paper<sup>42</sup> has been used in the present work to describe the catalytic metal center in the PMHs.

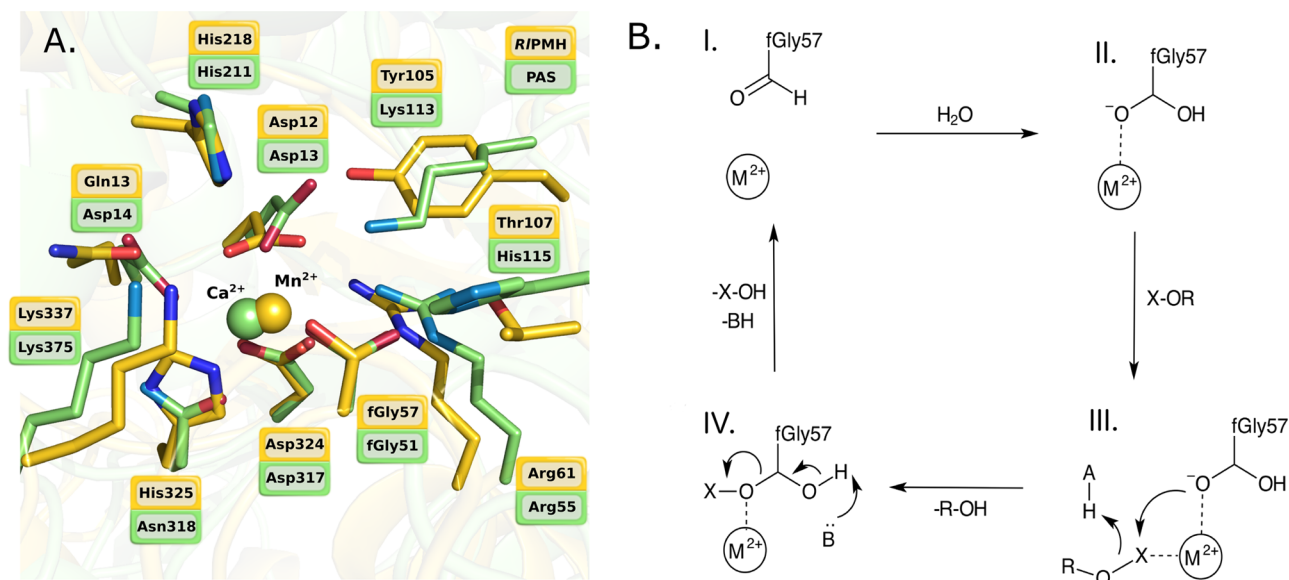
All relevant reactions were simulated in the active sites of both enzyme species and in a 24 Å water droplet, in order to quantify the catalytic effect of the enzyme compared to background reaction in aqueous solution. To model the reaction in solution, we used truncated residues to model the nucleophile (acetaldehyde hydrate as a model for the formylglycine) and the relevant general acids (ethylamine and ethylimidazole for Lys and His, respectively). In the enzyme simulations, one of the main computational difficulties encountered comes from the fact that truncating the 16 C-terminal residues of *RIPMH* causes the enzyme to lose its tetrameric structure, with a corresponding loss of activity. This strongly suggests that interactions at the subunit interfaces can be important to catalysis, as can also be observed from the protrusion of the interfacial loop almost

into the active site of the adjacent subunit (see Figure 2). Thus, it is necessary to include the entire (2056 amino acid) tetramer, which creates substantial computational cost.

To simplify this problem and reduce computational cost, the system was divided into three layers: the EVB (reacting) atoms, an active region encompassing all residues within a 24 Å sphere of the reacting atoms centered at the metal center, and an external layer in which the remainder of the system was present, but the atoms were constrained to their crystallographic positions (as is commonly done in similar studies, see e.g., refs 24 and 45). The simulation sphere encompassing the active region was centered on the catalytic Mn<sup>2+</sup> ion, and all crystallographic water molecules within 18 Å of this center were retained in our simulations, with the exception of any crystal waters clashing directly with the substrate once it was placed in the active site. The solvation sphere was then completed and extended to 24 Å using TIP3P<sup>46</sup> water molecules subjected to the surface constraint all atom solvent (SCAAS) spherical boundary conditions.<sup>47</sup> A 10 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> harmonic restraint was applied to the outer layer of the active region and associated solvent molecules (15%, 3.6 Å), in order to ease the transition between the active and constrained regions, which is why only an 18 Å of crystallographic water molecules were retained for the simulation. All forces on the constrained atoms were set to zero, in order to prevent them from distorting the dynamics of the active region. Ionizable residues within ~18 Å of the center were ionized during the course of the simulation, leading to a total system charge of -1 (without including the substrate). The protonation states of histidine side chains were investigated using the MolProbity server,<sup>41</sup> PROPKA 3.1,<sup>48,49</sup> and by visual inspection. All other residues, in particular those outside the active region, were set to their neutral form for system stability. The Mn<sup>2+</sup> ions in the adjacent monomeric units (the positions of which were kept constrained) were removed in order to avoid the presence of residual charge outside the simulation sphere (we note that the adjacent active sites all fall within the constrained external layer and all surrounding residues are therefore not allowed to move). In contrast, the catalytic metal center in the active region was described using a 7-pointed dummy model with distributed charges as described above.

**Molecular Dynamics Equilibration of the Systems of Interest.** All molecular dynamics (MD) and EVB simulations in this work were conducted using the OPLS-AA force field<sup>50</sup> implemented in the Q simulation package (Version 5.0.6).<sup>51</sup> For the substrate and nucleophile, OPLS-AA compatible force field parameters were generated with MacroModel 9.1 (OPLS-AA force field, 2001, Schrödinger LLC).<sup>52</sup> The only exceptions were the force field parameters for the carbon and oxygen of the deprotonated geminal diol, which were available in the literature and obtained from ref 53. Partial charges for the reacting atoms were generated at the HF/6-31G\* level of theory using the Gaussian 09 simulation package,<sup>54</sup> followed by the standard RESP procedure.<sup>55</sup>

All the simulations performed herein used time steps of 1 fs, while the temperatures of the system were regulated using the Berendsen thermostat<sup>56</sup> (with a 100 fs bath relaxation time). The systems were initially heated from 1 to 300 K over a short 80 ps simulation, applying a 200 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> harmonic force constant on the solute atoms to restrain them to their crystallographic positions. This allowed for the solvent molecules to equilibrate around the protein and the removal of initial contacts due to substrate placement. The system was then cooled down to 5 K for another 10 ps and then gradually heated to 300 K for 90 ps of simulation time, while the force constants of the harmonic restraint were gradually decreased from 200 to 0.5 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. Subsequently, a 5 ns equilibration was performed at 300 K for both wild-type and mutant enzyme simulations (300 ps for the reference reaction in solution) using a 0.5 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> position restraint on the substrate atoms, the side chain of the nucleophile, the catalytic metal center, and the side chain of the general acid (H218 or K337, depending on the mechanism being considered) to keep the reacting atoms in place. An RMSD plot of the active monomer for the wild-type enzyme and each enzyme variant is shown in Figures S2 and S3. As shown in this figure, due to the fixed excluded region, these systems equilibrated rapidly, with RMSD of < 0.5 Å from the reference



**Figure 3.** (A) Overlay of the active sites in R/PMH and PAS, illustrating conservation of active site structure between the two enzymes. (B) A simplified version of the proposed catalytic mechanism for both PMHs considered in this work, based on refs 8 and 12. Since the catalytic metal is suggested to be Mn<sup>2+</sup>, which is a hard Lewis acid, the nucleophile could be stable as an alkoxide, in agreement with the pH-rate profiles shown in Figure S4.

crystal structure. After the final equilibration step, we ran an additional 500 ps of molecular dynamics, during which 10 snapshots of the whole system were taken every 50 ps to be used as starting points for subsequent EVB simulations. Finally, although we remained as faithful as possible to a fully nonbonded model for the catalytic metal center, we introduced an angle parameter (50 kcal·mol<sup>-1</sup>·rad<sup>-2</sup>, equilibrium angle 180°) between the center of the Mn<sup>2+</sup> dummy model and the residue D324 (Mn<sup>2+</sup>–O<sub>metal</sub>–O<sub>free</sub>, where O<sub>metal</sub> corresponds to the oxygen atom closest to the Mn<sup>2+</sup>, and O<sub>free</sub> the one not coordinated to it), which would otherwise become bidentately coordinate to the metal center and make the active site unstable. As can be seen from the Results and Discussion, despite the inclusion of this extra parameter, we are able to systematically reproduce the activation energies of both wild-type and mutant forms of these PMHs with different substrates with good agreement to experimental data.

**Empirical Valence Bond Calculations.** Our methodology of choice in this work to model chemical reactivity was the EVB approach of Warshel and co-workers.<sup>32,33</sup> This is an empirically-based multiscale valence-bond/molecular mechanics approach that is fast enough to allow for the extensive sampling required to obtain convergent free energies for complex biochemical processes, while having a proven track record as a powerful tool for quantifying and rationalizing the catalytic power of native and mutant enzymes.<sup>36,44</sup> All EVB calculations were performed using the standard EVB free energy perturbation/umbrella sampling (EVB-FEP/US) procedure outlined in refs 33 and 57, as implemented in the Q simulation package.<sup>51</sup>

The reaction under study was described in terms of two valence bond structures, as illustrated in Section S4 of the Supporting Information. It should be pointed out that as all atoms in the two valence bond states are treated using the same force field, the only differences between the reacting (EVB) and nonreacting atoms are the use of Morse rather than harmonic potentials to describe bonds that are being broken and formed during the reaction (see the Supporting Information) and the fact that unlike the rest of the protein, the EVB atoms do not have a cutoff for calculation of the nonbonded interactions. All EVB-FEP/US simulations were performed at 300 K, using 51 mapping windows of 200 ps per window, resulting in 10.2 ns of simulation time for each individual trajectory, sampling over 10 starting conformations per system (102 ns per system) and ~4 μs cumulative simulation time for all systems studied in this work. All MD and EVB simulations were performed using a 1 fs time step, and long-

range effects were treated using the local reaction field (LRF) approach.<sup>58</sup>

Finally, as outlined above, we also modeled the corresponding uncatalyzed reaction for each substrate of interest in this work, as we needed these calculations not only for the calibration of the EVB parameters, but also to compare the reactions for different environments. Coordinates for each system were based on the equilibrated enzyme system, but in the absence of the enzyme itself, and the reactions were modeled using the relevant substrate as well as acetaldehyde and protonated ethylamine as models for the nucleophile and the general acid, as described in the Initial System Setup section. All equilibration and EVB protocols were the same as for the full enzyme system, with the exception that the background reaction in the absence of the enzyme was only equilibrated for 300 ps rather than 5 ns. As with the corresponding enzymatic reactions, a weak position restraint of 0.5 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> was placed on all solute atoms (nucleophile, substrate, and model for general acid) in order to keep the reacting fragments in the center of the simulation sphere. This weak restraint is sufficient to keep the nitrogen atom of the general acid within 3.5 Å of the leaving group oxygen throughout the 300 ps equilibration of the background reaction, in part due to electrostatic interactions between the charge on the general acid and the substrate. The relevant background reaction was calibrated based on estimations using experimental data, as described in detail in the Supporting Information. Furthermore, the parameters describing the relative positions of the VB parabolas and the coupling between them were then transferred *unchanged* to the enzyme in order to be able to quantify and correctly predict the catalytic effect of wild-type and mutant enzymes.

## RESULTS AND DISCUSSION

**Modeling the Catalytic Mechanism for the Wild-Type Enzymes.** A unique feature of the PMHs being considered in the present work is that they are the only nonsulfatase enzymes known to date to possess a post-translational modification from a cysteine to an aldehyde (formylglycine, fGly).<sup>8,12,19</sup> The current proposed mechanism<sup>8</sup> for both native and promiscuous PMH activities is shown in Figure 3. In a first step (I → II), hydration of the post-translationally modified aldehyde yields a reactive geminal diol, which can act as a nucleophile. This

geminal diol is activated by the catalytic metal center and exists in its alkoxide form. Following substrate binding (**II**  $\rightarrow$  **III**), this geminal diol then attacks the phosphorus/sulfur center (**III**  $\rightarrow$  **IV**) of the relevant substrate (Figure 2) to give rise to a hemiacetal intermediate (**IV**). In the final step (**IV**  $\rightarrow$  **I**), this intermediate is hydrolyzed by hemiacetal cleavage to regenerate the aldehyde and yield the final product.

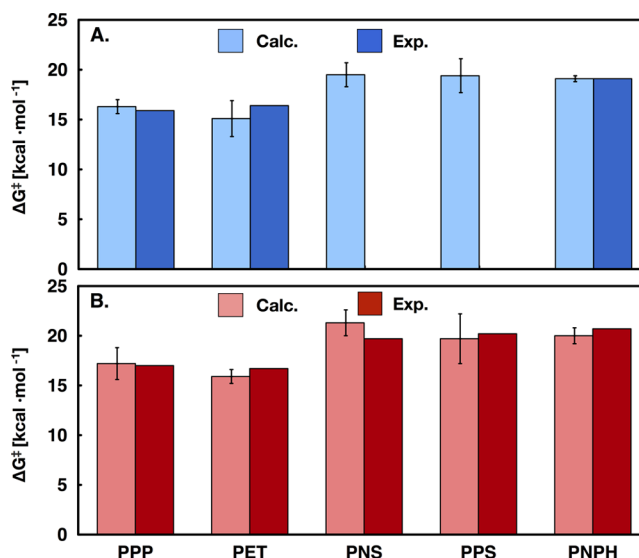
The two PMHs considered in this work show an absolute dependence on divalent metal ions, with the most likely candidate for fulfilling this role being  $\text{Mn}^{2+}$ , based on experimental data presented in refs 8 and 12. A transition metal would be expected to substantially decrease the  $\text{p}K_{\text{s}}$  of the metal-bound nucleophile to yield an alkoxide, as also suggested by the acid limb of the pH-rate profiles shown in Figure 2 of ref 12 (reproduced as Figure S4), which most likely corresponds to the deprotonated nucleophile (see discussion in ref 12). The pH-rate profiles, which are coincidental for all substrates except the phosphate triester, also suggest the involvement of an acid catalyst, most likely either H218 or K337<sup>12</sup> (see also Figures 3 and S4).

It has been argued that steps **IV**  $\rightarrow$  **I** of Figure 3 can play an important role in facilitating promiscuity.<sup>12</sup> Specifically, harnessing hemiacetal cleavage allows for a common mechanism irrespective of the functional group used in the intermediate,<sup>12</sup> while simultaneously providing a thermodynamically less challenging route to facilitate C–O cleavage, compared to the repeated cleavage of an extremely stable P(S)–O bond.<sup>59</sup> However, kinetic data on base-catalyzed hemiacetal cleavage in aqueous solution demonstrate that this reaction is extremely fast.<sup>60</sup> Additionally, as all substrates will be broken down by a common mechanism through a common intermediate, the selectivity will be already determined in steps **III**  $\rightarrow$  **IV**, which is also therefore the focus of the present work (Figure 3).

Our mechanistic model assumes an anionic nucleophile and general acid catalysis from either K337 or H218 to protonate the departing leaving group. As discussed below, in this work K337 was chosen as the general acid based on empirical  $\text{p}K_{\text{a}}$  calculations and experimental data. That is, the experimental pH-rate profiles<sup>12</sup> suggest a two- $\text{p}K_{\text{E}}$  model, with a  $\text{p}K_{\text{E}1}$  of 7.0–7.2 (5.8 for the sulfate monoester) and a  $\text{p}K_{\text{E}2}$  of 7.5–8.1 (see Figure S4). The first  $\text{p}K_{\text{a}}$  is likely to correspond to the nucleophile, contributing to catalysis in its deprotonated form as discussed above, and the second  $\text{p}K_{\text{a}}$  to the general acid. The  $\text{p}K_{\text{E}2}$ , which is very close to the  $\text{p}K_{\text{a}}$  of around 8 suggested for K337 by PROPKA, led to the choice of this residue as the putative general acid, as also suggested by refs 8 and 12. The only exception to this model is the *p*-nitrophenyl phosphate monoester, of which the dianionic form will be extremely resistant to attack by an anionic nucleophile (see ref 62). However, the  $\text{p}K_{\text{a}}$  of the already basic nonbridging oxygens of this substrate ( $\text{p}K_{\text{a}} \sim 5.0$ <sup>63</sup>) is likely to be substantially elevated due to the close proximity of the anionic nucleophile. This, in turn raises the possibility that this substrate binds as a monoanion, as has already been demonstrated by simulations, for example, for protein tyrosine phosphatase 1B.<sup>64</sup> Note also that, as shown in Figure S4 and ref 12, the monoanionic sulfate and dianionic phosphate monoesters give rise to very different pH-rate profiles that not only have different slopes but also are shifted by 2 pH units. Thus, we have herein considered a mechanism involving an anionic nucleophile attacking a monoanion phosphate, which yields excellent agreement with experiment as discussed below.

In the present work we have not examined phosphate triester hydrolysis, which shows a very different  $k_{\text{cat}}/K_{\text{M}}$  pH-rate profile to all other substrates studied (Figure S4). The inverted pH-rate profile observed for this substrate suggests either the involvement of a completely different set of residues or a completely different mechanism of catalysis. Additionally, BcPMH shows extremely poor activity toward this substrate ( $k_{\text{cat}}/K_{\text{M}}$  of  $1.6 \times 10^{-2} \text{ M}^{-1}\cdot\text{s}^{-1}$ ), which is actually slower than the corresponding uncatalyzed alkaline hydrolysis of the model substrate paraoxon.<sup>65</sup> Taken together, this suggests that the hydrolytic mechanism of this substrate, if it at all binds in the same active site, is impossible to prove conclusively through calculations due to lack of concrete experimental data.

As mentioned before, the reactions examined in this work correspond to the third step (**III**  $\rightarrow$  **IV**) of the catalytic cycle shown in Figure 3. Since experiments show that the hemiacetal cleavage is a fast step, we focused only on this second step, which is the most chemically challenging step of the cycle, being thus the one related to the measured kinetic parameters. Figure 4 shows a comparison between our calculated and,



**Figure 4.** Calculated and experimentally derived activation energies for the enzyme-catalyzed reactions of the five substrates studied here by the wild-type forms of (A) R1PMH and (B) BcPMH.<sup>61</sup>

where available, experimental activation free energies (derived from  $k_{\text{cat}}$ , which provides an upper limit for the reaction rate).<sup>8,12</sup> The corresponding tabulated values can be found in Table 1. From our results, it can be seen that the model used in the present work reproduces the experimental activation free energies within an accuracy of 1.7 kcal·mol<sup>-1</sup> for all substrates. It has additionally been argued<sup>8</sup> that the PMHs considered in this work can accept both diesters and phosphonates with such high proficiency in the same active site due to similar geometrical and steric demands for the respective substrates and transition states. To probe this further, we have examined transition-state geometries for all uncatalyzed and enzyme-catalyzed reactions considered in this work. Table 2 shows a comparison of P(S)–O distances to the oxygen atoms of the incoming nucleophile and departing leaving group for all substrates and reactions. Representative transition-state structures in the BcPMH active site are also illustrated in Figure 5. From these results, it can be seen that the PMHs hydrolyze all substrates through a unified mechanism with similar substrate

**Table 1.** Calculated<sup>a</sup> and Observed<sup>b</sup> Activation ( $\Delta G^\ddagger$ ) and Reaction Free Energies ( $\Delta G^\circ$ ) for the Hydrolysis of the Five Substrates by the Wild-Type Forms of *R/PMH* and *BcPMH*<sup>c</sup>

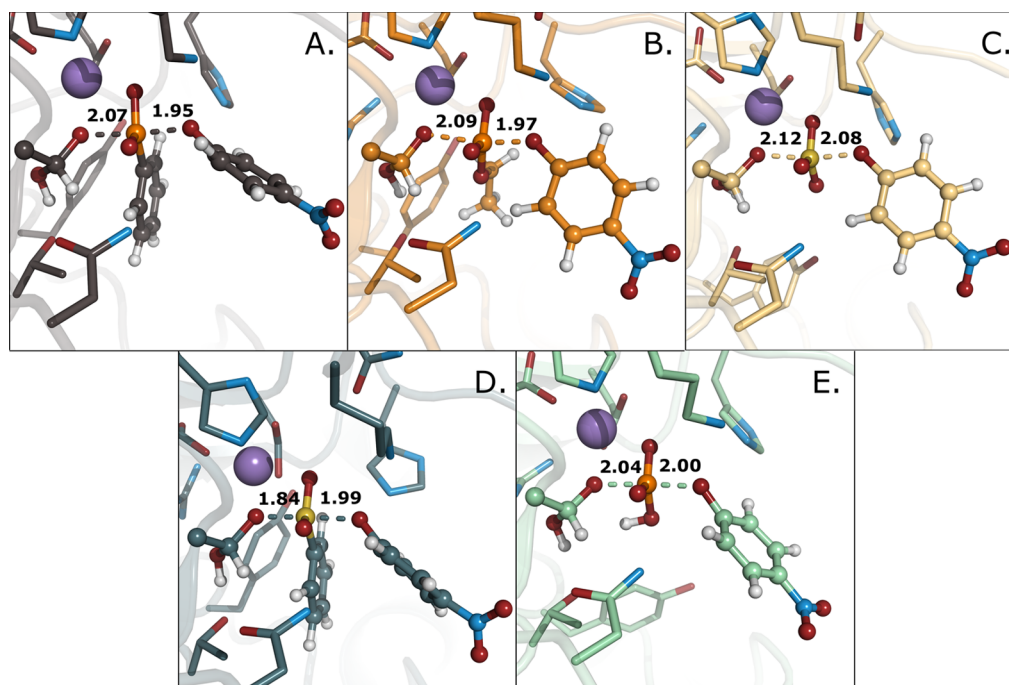
substrate	$\Delta G^\ddagger$		$\Delta G^\circ$		$\Delta G^\ddagger$		$\Delta G^\circ$
	calc.	expt.	calc.	calc.	expt.	expt.	calc.
						<i>R/PMH</i>	<i>BcPMH</i>
PPP	16.3 ± 0.7	15.9	-16.9 ± 1.1	17.2 ± 1.6	17.0		-15.7 ± 2.8
PET	15.1 ± 1.8	16.4	-6.3 ± 2.0	15.9 ± 0.7	16.7		-6.3 ± 1.1
PNS	19.5 ± 1.2	n.d.	-2.0 ± 1.5	21.3 ± 1.3	19.7		-6.2 ± 1.9
PPS	19.4 ± 1.7	n.d.	1.3 ± 2.1	19.7 ± 2.5	20.2		3.2 ± 3.6
PNPH	19.1 ± 0.3	19.1	-9.2 ± 1.0	20.0 ± 0.8	20.7		-7.8 ± 1.2

<sup>a</sup>“Expt.” and “calc.” denote experimental and calculated values respectively, and “n.d.” refers to values that have not been experimentally determined. All energies are given in kcal·mol<sup>-1</sup> and are averages and standard deviations based on 10 individual EVB simulations generated from different starting structures, as outlined in the Methodology section. <sup>b</sup> $\Delta G^\ddagger$ (expt.) corresponds to experimental values of the enzyme-catalyzed reaction, based on the kinetic data presented in refs 8 and 12. <sup>c</sup>The corresponding EVB parameters are presented in the Supporting Information.

**Table 2.** Average P(S)–O Distances, in Å, at the Transition State for the Relevant Group Transfer Reaction in Water and in the Wild-Type forms of *R/PMH* and *BcPMH*<sup>a</sup>

	water		<i>R/PMH</i>		difference <i>R/PMH</i> – water		<i>BcPMH</i>		difference <i>BcPMH</i> – water	
	P/S–O <sub>nuc</sub>	P/S–O <sub>lg</sub>	P/S–O <sub>nuc</sub>	P/S–O <sub>lg</sub>	P/S–O <sub>nuc</sub>	P/S–O <sub>lg</sub>	P/S–O <sub>nuc</sub>	P/S–O <sub>lg</sub>	P/S–O <sub>nuc</sub>	P/S–O <sub>lg</sub>
PPP	1.89 ± 0.08	2.14 ± 0.12	2.09 ± 0.11	1.96 ± 0.09	0.20	-0.18	2.07 ± 0.11	1.95 ± 0.08	0.18	-0.19
PET	2.09 ± 0.29	2.04 ± 0.12	2.06 ± 0.11	1.94 ± 0.08	-0.03	-0.10	2.09 ± 0.13	1.97 ± 0.09	0.00	-0.10
PNS	2.13 ± 0.17	2.06 ± 0.13	2.11 ± 0.13	2.08 ± 0.12	-0.02	0.02	2.12 ± 0.13	2.08 ± 0.12	-0.01	0.02
PPS	1.85 ± 0.07	2.03 ± 0.09	1.88 ± 0.07	1.99 ± 0.08	0.03	-0.01	1.84 ± 0.06	1.99 ± 0.08	0.03	-0.01
PNPH	1.99 ± 0.25	2.14 ± 0.18	2.03 ± 0.11	2.0 ± 0.11	0.04	-0.14	2.03 ± 0.10	2.03 ± 0.10	0.04	-0.11

<sup>a</sup>All data are averages and standard deviations over 10 individual simulations as outlined in the Methodology section. For an extended version of this table, including both P(S)–O and H–N/O<sub>lg</sub> distances, we refer the reader to the Table S4.

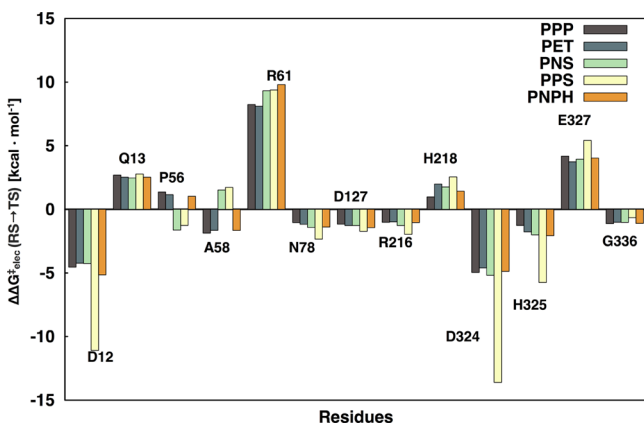
**Figure 5.** Representative transition-state structures for the *BcPMH* catalyzed hydrolysis of (A) PPP, (B) PET, (C) PNS, (D) PPS, and (E) PNPH. Values presented here correspond to the averages over 10 trajectories.

binding positions and transition states. With the exception of the phosphonate, little change is seen in transition-state geometry upon moving from aqueous solution to the enzyme active sites, in agreement with related experimental work by Herschlag and co-workers,<sup>22</sup> as well as theoretical analysis by Hou and Cui,<sup>25</sup> on alkaline phosphatase. Even in the case of the

phosphonate, the overall transition-state size (considering the distance between O<sub>nuc</sub>–O<sub>lg</sub>) stays very similar, and the main change is that the symmetry of the transition states changes, with P–O<sub>nuc</sub> becoming slightly elongated and P–O<sub>lg</sub> slightly compressed compared to the corresponding uncatalyzed reaction. Hence, as suggested in previous works<sup>22,25</sup> for alkaline

phosphatase, we find very little effect on the transition-state geometries of moving to the enzyme active when compared with those obtained through modeling the corresponding uncatalyzed reaction in aqueous solution.

To locate the origin for the differences in the observed catalytic activity, we have performed a comparison of the electrostatic contributions of individual residues to the hydrolysis of each substrate, calculated using the linear response approximation following previous works (e.g., refs 66 and 67; see Figure 6). This analysis shows that, strikingly, despite the



**Figure 6.** Electrostatic contribution of key residues to the calculated activation barrier for the hydrolysis of the five substrates for wild-type form of BcPMH.<sup>72</sup>

calculations of residue interactions being completely independent of each other, with different charge distributions and different transition states, all residues that make substantial electrostatic contributions to the calculated activation barrier are conserved among different substrate. This is similar to the observations from our previous computational work on the arylsulfatase from PAS.<sup>13</sup> However, although qualitatively similar, there are some key quantitative differences between the different substrates, most notably in the case of D12, D324, H325, and E327. While this itself is hardly surprising, considering these are electrostatically quite different substrates, it highlights that the active site pre-organization is not “perfect”, but rather cooperative electrostatic interactions render this preorganization flexible enough to readily adapt to the electrostatic needs of different substrates (see also the related discussion of catalytic backups in serum paraoxonase 1).<sup>68</sup>

To further explore this observation, we have also examined the charge change on the central P/S atom and all atoms bonded to it upon moving from reactant to transition state for the wild-type BcPMH catalyzed hydrolysis of the different substrates studied in the present work. These atoms were split into three fragments: a central fragment comprising the P/S atom, the nonbridging oxygens of the substrate, and the C atom of the phenyl group of phenyl *p*-nitrophenyl sulfonate (PPS) and phenyl *p*-nitrophenyl phosphonate (PPP) connected to the central P/S atom as well as the oxygen atoms of the departing leaving group ( $O_{lg}$ ) and attacking nucleophile ( $O_{nuc}$ ) as individual fragments (see Table S3). As the transition states for the reactions studied involve partial bond formation to the nucleophile and leaving group, we have summed up the charges on the central atoms (P/S, nonbridging oxygens, and carbon) and treat this as one unit, which we will henceforth refer to as “central fragment” for simplicity. The schematic for this division

is shown in Table S3 which also provides absolute charges for each fragment at the reactant and transition states as well as the charge shift upon moving from reactant to transition state. These have then in turn been ranked against the measured  $k_{cat}/K_M$  for BcPMH for each substrate.<sup>12</sup> From this table, it can be seen that while there is little trend in the charge shift on the leaving group oxygen (which is partially protonated by the general acid), there are subtle but clear trends in the partial charges of the nucleophile oxygen and the central fragment. That is, for the native substrate, PPP, there is a substantial charge shift corresponding to a loss of +0.2728 au on  $O_{nuc}$ . This charge shift gradually decreases across the series, correlated with a reduction in  $k_{cat}/K_M$ . In parallel to this, for the native substrate, there is a small buildup of negative charge (−0.0468 au) on the central fragment, which increases to −0.0884 au for the promiscuous activity with the lowest observed  $k_{cat}/K_M$  (the sulfate monoester). This suggests a subtle preference for minimizing negative charge at the transition state. Therefore, there appears to be a correlation between the calculated charge shift at the transition state and the subsequent catalytic efficiency of the enzyme. This also ties in with experimental observations that other alkaline phosphatases such as AP and NPP clearly discriminate on the basis of substrate charge.<sup>69–71</sup> The most radical example of such charge discrimination in this superfamily, in fact, is in the arylsulfatase from *Pseudomonas aeruginosa* (PAS), where the hydrolysis of large bulky monanionic diesters such as bis-*p*-nitrophenyl phosphate shows only 100-fold lower values than the monanionic substrate *p*-nitrophenyl sulfate (with  $k_{cat}/K_M = 4.9 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$  for the sulfate monoester and  $2.5 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$  for the phosphate diester).<sup>9</sup> In contrast, the dianionic analogue of the sulfate monoester, *p*-nitrophenyl phosphate, is a much poorer substrate than the sulfate monoester ( $k_{cat}/K_M = 790 \text{ M}^{-1} \cdot \text{s}^{-1}$ ),<sup>9</sup> despite having the same ground-state geometry and similar predicted transition-state geometries to the sulfate monoester. Even further examples of such charge discrimination have been seen by Baxter and co-workers<sup>73</sup> in studies of aluminum and magnesium fluoride transition-state analogues (TSA) of phosphoryl transfer enzymes, where they showed clear preference for preserving anionic charge at the expense of TSA geometry over a broad range of pH.

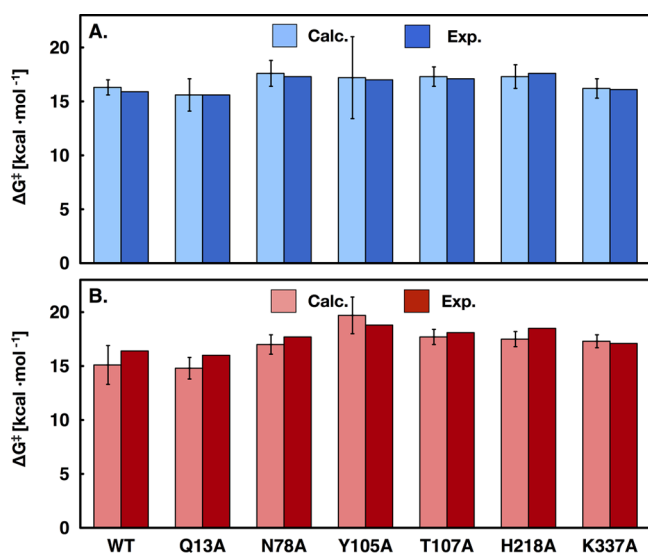
Clearly, these enzymes have evolved to provide the key active site interactions that optimally stabilize the transition state for the native reaction, which in turn leads to the observed preference for the native substrate. However, in the electrostatic cooperativity model we present in this work, these interactions are sufficiently flexible to accommodate the electrostatic needs of other substrates, although the same residues can make quantitatively different contributions as shown in Figure 6. These differences in ability to stabilize different transition states would in turn lead to the selectivity displayed by these enzymes for different substrates.

**Probing Key Active Site Mutations.** Hollfelder and co-workers<sup>8</sup> have performed a detailed alanine scan of the R1PMH active site residues, testing against both the phosphonate (PPP) and phosphodiesterase (PET) activities of the enzyme. Both substrates appear to be highly insensitive to active site single mutations, with the individual substitution of each key active site residue in R1PMH leading to, at worst, a ~20-fold reduction in  $k_{cat}$ . An exception to this is modifying the nucleophile to alanine, but even in this extreme case, these enzymes still show some activity.<sup>8,12</sup> To probe the origin of the seeming resilience of these enzymes to substitution of



individual active site residues, we have performed EVB calculations to obtain free energy profiles for the chemical step for the hydrolysis of substrates PPP and PET of Figure 2 by a range of Ala-substituted forms of *RIPMH* presented in ref 8. Critically, when moving from wild-type to mutants, we used *exactly* the same parameter set, unchanged, allowing us to in parallel rigorously validate our valence bond model for the reaction mechanism catalyzed by these enzymes and its predictive power (see the Supporting Information for theoretical background). The only exception to this is the K337A mutant that has the general acid functionality of K337 removed, and for which we model the reaction as proceeding with H218 as the general acid instead (for all other calculations, H218 is kept in its neutral form as close proximity to the K337 side chain and the catalytic metal ion will depress its  $pK_a$ ). Note again that although we have focused on K337 as a putative general acid in this work, due to the agreement between the predicted  $pK_a$  of K337 and the experimental pH-rate profiles as outlined in the previous section, in practice either of these two residues could fulfill the role of general acid.

A comparison between calculated and experimental activation barriers for the hydrolysis of these substrates by each key *RIPMH* variant is shown in Figure 7 with the



**Figure 7.** Calculated and experimentally derived activation energies for the *RIPMH* catalyzed reactions of (A) PPP and (B) PET. Shown here is data both for the reaction catalyzed by the wild-type enzyme as well as several mutant forms of the enzyme. The data plotted in this figure are presented in Table 3, and the error bars represent standard deviations over 10 independent trajectories.

corresponding energetics presented in Table 3. It can be seen that we are able to reproduce the experimentally observed effect of all active site mutants ( $\Delta\Delta G^\ddagger_{WT \rightarrow mut}$ ) to within an average error margin of  $\sim 1$  kcal·mol<sup>-1</sup>. The most challenging of these mutations to model is the Y105A mutation (as can also be seen from the large error bar shown in Figure 7), as this mutation leads to a larger perturbation in the active site, for example repositioning residues such as T107 and Y215. This results in a larger standard deviation for these calculations in the case of the phenyl phosphonate than in other simulations. However, the average over 10 trajectories is still in good agreement with experimental results. Taken together with our other simulations, this provides support for the quality of our calculations,

the suggested mechanism, and our assumption that the group transfer is the key step in determining the specificity. Additionally, an examination of the corresponding P–O distances at the transition state for each variant and substrate shows that, as we move from the background reaction to the enzyme (Table S5), the single mutants in the active site have little effect on the transition-state geometry.

Here, we will provide a brief discussion of our simulations of each mutant below and refer the reader to Figure 5 for an overview of how each residue interacts with the substrates of interest in the equilibrated wild-type enzyme.

**Q13A.** As suggested in ref 8, this residue seems to play a key role by holding K337 in place. For the simulations of the Q13A variant, we observe that both K337 and H218 are perturbed, and the RMS displacement of K337 after equilibration compared to the wild-type enzyme is 0.80 and 0.78 Å for PPP and PET, respectively. However, the effect of losing this interaction translates to only a 0.4 kcal·mol<sup>-1</sup> reduction in activation barrier for both substrates both experimentally and from our simulations (see Table 3). Note that, as shown in Table S2, this mutation primarily affects  $K_M$  rather than  $k_{cat}$  for both substrates considered here.

**N78A.** N78 is a key active site residue, as it provides a hydrogen-bonding interaction to the nonbridging oxygen of both PPP and PET. This interaction stabilizes the substrate and also helps to optimally position the substrate in the active site. Loss of this interaction results in a 1.4 kcal·mol<sup>-1</sup> increase in barrier for both substrates. As shown in Table 3 and Figure 7 we are able to reproduce the detrimental effect of the N78A mutant, which is slightly larger for PET than for PPP. This could be in part due to the presence of the phenyl ring in PPP, which can help to position the substrate in the active site even in the absence of the hydrogen bond from N78.

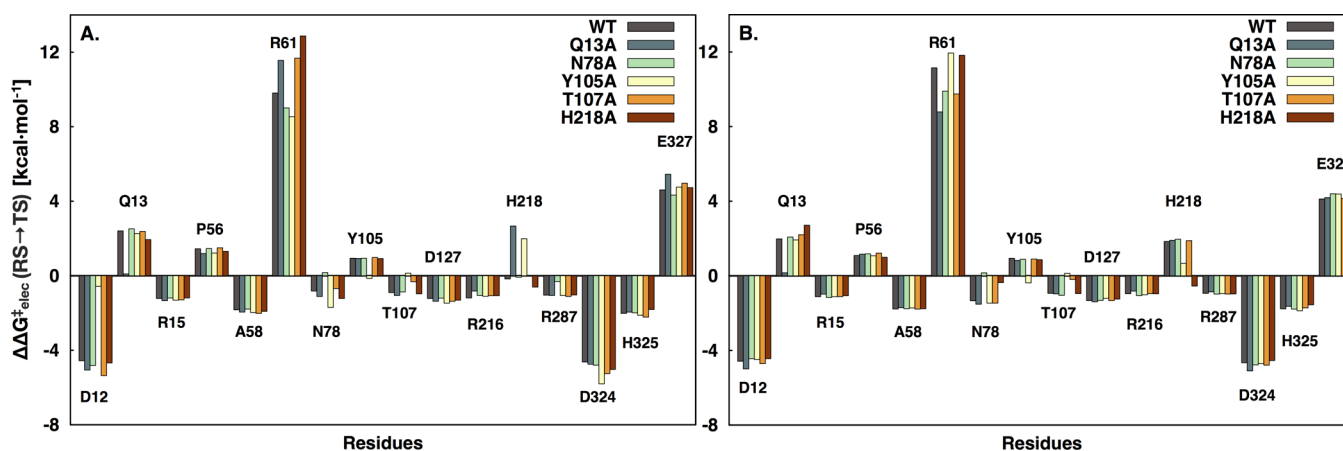
**Y105A.** In the wild-type enzyme this residue does not directly interact with the nucleophile or the substrate. However, it is part of a hydrogen-bond network that keeps D324 and R61 correctly positioned for catalysis (Figure 1). It has been experimentally shown that this mutation drastically reduced the kinetic efficiency of the enzyme. As can be seen from Table 2, while we are able to reproduce the experimental activation barrier within 1 kcal·mol<sup>-1</sup>, for both the phosphonate monoester and the phosphate diester, we obtain larger standard deviations for this variant in the case of the phosphonate monoester. In the case of the phosphate diester, upon truncating Y105 to alanine and equilibrating the system, we see an increase in the number of water molecules around the nucleophile as well as repositioning of other residues, such as T107 and Y215. Specifically, the interaction between the nucleophile and T107 is broken, and Y215 occupies the space left by mutation. In the case of the phosphonate monoester, however, the large hydrophobic phenyl ring of the phosphonate (compared to the smaller ethyl group in the diester) blocks water access to the nucleophile and restricts the movement of other residues around it.

**T107A.** T107 is another key active site residue, as it directly interacts with the nucleophile, helping optimally position it for catalysis (Figure 1). Unsurprisingly, truncating this residue to alanine leads to an increase in activation barrier of 1.2 kcal·mol<sup>-1</sup> for the phosphonate and 1.7 kcal·mol<sup>-1</sup> for the diester (a trend we reproduce computationally, see Table 3 and Figure 7). This is primarily due to loss of the hydrogen bonding interaction with T107 as well as the resulting subtle repositioning of the nucleophile in the active site.

**Table 3.** Calculated<sup>a</sup> and Observed<sup>b</sup> Activation ( $\Delta G^\ddagger$ ) and Reaction Free Energies ( $\Delta G^\circ$ ) for the Hydrolysis of PPP and PET in Both the Wild-Type and Different Mutant Forms of *RIPMH*<sup>c</sup>

system	$\Delta G^\ddagger$		$\Delta G^\circ$	$\Delta G^\ddagger$		$\Delta G^\circ$
	expt.	calc.	calc.	expt.	calc.	calc.
	PPP			PET		
WT	15.9	16.3 ± 0.7	-16.9 ± 1.1	16.4	15.1 ± 1.8	-6.3 ± 2.0
Q13A	15.6	15.6 ± 1.5	-18.0 ± 1.4	16.0	14.8 ± 1.0	-10.5 ± 1.4
N78A	17.3	17.6 ± 1.2	-18.4 ± 1.7	17.7	17.0 ± 0.9	-5.8 ± 2.0
Y105A	17.0	17.2 ± 3.8	-13.8 ± 3.6	18.8	19.7 ± 1.7	-5.6 ± 1.9
T107A	17.1	17.3 ± 0.9	-9.9 ± 1.8	18.1	17.5 ± 0.7	-3.3 ± 1.8
H218A	17.6	17.3 ± 1.1	-13.4 ± 1.2	18.5	17.5 ± 0.7	-5.7 ± 1.2
K337A	16.1	16.2 ± 0.9	-7.1 ± 1.3	17.1	17.5 ± 0.8	-2.2 ± 1.9

<sup>a</sup>“Expt.” and “calc.” denote experimental and calculated values, respectively. All energies are given in kcal·mol<sup>-1</sup> and are averages and standard deviations over 10 individual trajectories using different starting conformations, as outlined in the Methodology section. <sup>b</sup> $\Delta G^\ddagger$ (expt.) corresponds to experimental values of the enzyme-catalyzed reaction, based on the kinetic data presented in refs 8 and 12. <sup>c</sup>The corresponding EVB parameters are presented in the Supporting Information.



**Figure 8.** Electrostatic contribution of key residues to the calculated activation barrier for group transfer reactions of (A) PPP and (B) PET for different *RIPMH* variants.

*H218A*. Interestingly, this mutation appears not only to impact the catalytic activity, but also to increase  $K_M$  from 2.8 mM in the wild-type enzyme to 15 and 57 mM in the mutant form for both phosphonate and diester substrates.<sup>8</sup> It has been argued that this increase in  $K_M$  is due to substrate binding, suggesting that this residue as well as K337 are directly involved in this step. H218 could also play a role as a general acid, due to its close proximity with the leaving group. However, in its unprotonated form, H218 also plays a key role in positioning K337 for optimal leaving group stabilization (the distance between  $N_e$  of H218 and K337 is 3.44 Å in the *RIPMH* crystal structure).

*K337A*. The role of K337 is two-fold: it helps to position the substrate in the active site in the Michaelis complex (through a hydrogen bonding interaction to one of the nonbridging oxygens of the substrate) as well as to stabilize the leaving group upon departure by acting as a general acid and protonating it. One would expect, then, that mutation of this key residue to alanine would result in a substantial increase in activation barrier. However, the experimentally observed increase is only 0.2 kcal·mol<sup>-1</sup> for the phosphonate monoester and 0.7 kcal·mol<sup>-1</sup> for the phosphate diester, which is lower than for example either the N78A or T107A mutations. This suggests that another positively charged residue is taking up the role of K337 in leaving group stabilization. As discussed above, the role of general acid could be fulfilled by either K337 or

H218, and in absence of the close proximity of the K337 positive charge upon mutation (the distance between the  $N_e$  of H218 and the nitrogen of K337 is 3.44 Å in the wild-type crystal structure),<sup>8</sup> one would presume that H218 is more likely to be protonated than in the wild-type. Therefore, we tested modeling H218 as a general acid, demonstrating that this in fact provides activation barriers in very good agreement with experiment (Table 3), suggesting that in the absence of K337, H218 takes up the role of this residue in leaving group stabilization (either through hydrogen-bonding/charge–charge interactions with the anionic leaving group or as a general acid).

So far, we have not yet discussed the details of the proton transfer to the leaving group from either K337 or H218. In the present work, we have used a two-state valence bond model to describe this process, as outlined in the Methodology section. In our model, for both the lysine- and histidine-catalyzed mechanisms (wild-type and K337A mutants, respectively), the group transfer and proton transfer reactions take place in a single, concerted but slightly asynchronous reaction step (Figures S5 and S6 and Table S4–S6). As seen from these figures and associated table, when the general acid is modeled as being K337, the transition state is dominated by the group transfer reaction, with the proton transfer reaction taking place very slightly after the group transfer. This would tie in with the fact that the *p*-nitrophenol leaving group is sufficiently basic to not *a priori* need protonation to depart. In the case of H218 as

the general acid (in the K337A mutant), the proton transfer becomes more concerted with the group transfer reaction (Figure S6). This would be in agreement with our previous DFT study of the hydrolysis of phosphate and sulfate monoesters,<sup>18</sup> in which we carefully examined all proton transfer steps involved and demonstrated that they either immediately precede or succeed the group transfer step, but along the same reaction coordinate (without the need for discrete intermediates). Such a model also agrees with high-level QM/MM calculations of the phosphoryl transfer reaction catalyzed by dUTPase,<sup>74</sup> which show a similar coupling between proton transfer and group transfer reactions, suggesting that a two-state VB model is adequate for capturing the key features of the relevant reaction mechanisms, and this is also borne out by the agreement with the experimental data.

Finally, a comparison of the electrostatic contributions of individual residues to catalysis for both substrates and all variants (Figure 8) shows that, as with the wild-type reactions, changes in activity correspond to cooperative electrostatic effects, where the active site residues are able to compensate the absence of key active site residues and stabilize the transition state of different substrates. This effect was also seen in our previous computational studies of the evolutionarily related PAS<sup>13</sup> and has also been alluded to in other recent works.<sup>38,75</sup> A similar phenomenon has been observed in experimental studies of serum paraoxonase 1,<sup>68</sup> suggesting that such electrostatic flexibility is a feature of multiple enzymes that catalyze phosphoryl transfer.

## ■ EXAMINING OTHER PLAUSIBLE CONTRIBUTIONS TO THE OBSERVED SELECTIVITY AND PROMISCUITY

Although our data strongly point toward electrostatic cooperativity as the origin for the observed selectivity and promiscuity among members of this superfamily, it is important to also examine other possible origins of this effect. Before proceeding further in this discussion, it is worth mentioning that there are several different ways to define this concept,<sup>76</sup> that range from an enzyme performing distinct chemistry using a similar set of residues and the same mechanism to cases where different sites of an enzyme are used to perform different chemical reactions (a form of “protein moonlighting”). Common to all these definitions, however, is the fact that promiscuity can be regarded simply as a converse of specificity, in that a highly *specific* enzyme would only be able to perform a single chemical reaction, whereas a *catalytically promiscuous* enzyme would be able to perform multiple distinct chemical reactions.<sup>3,77</sup> In addition, while the enzymes studied in the present work are multifunctional, with very high proficiencies for both phosphonate monoester and phosphodiester hydrolysis, they nevertheless show high selectivity and an order of preference between these and other promiscuous reactions that they catalyze (see Table S1). Therefore, in the present discussion, we will use “specificity” as a converse to promiscuity (i.e., referring to the number of reactions the enzyme catalyzes) and “selectivity” to indicate the discrimination between different reactions catalyzed by the same enzyme.

An important point to take into account in the present work is that, with the exception of the monoesters, the reactivity of the substrates examined herein is substantially lower by several orders of magnitude under neutral conditions than at high pH (see Table S1). Therefore, it is plausible to consider that a part of the broad substrate specificity might result because the

enhanced reactivity of the active site nucleophile by the catalytic metal center already provides substantial rate accelerations for a broad range of substrates. This would be consistent with the small effects of  $< 10$ -fold (on  $k_{\text{cat}}$ ) on the enzymatic activity of the mutation of their key residues that interact with the substrate oxygens, specifically N78 and K337.<sup>8</sup> However, while having an activated nucleophile is clearly important for the overall activity toward different substrate, this in itself is not fully sufficient to describe the observed promiscuity, as there are clear variations in rate acceleration between the different substrates, even when considering the alkaline reaction as the relevant reference state for the uncatalyzed reaction (see values presented in Table S1).

Additionally, as discussed and demonstrated in several of our previous works,<sup>17,18,78</sup> despite the superficial similarities between the different transition states involved (substitution of P for S, adding or removing functional groups), they have very different charge distributions and thus solvation patterns, leading to very different requirements for efficient catalysis. This can also be seen in both the quantitative differences in the residue contributions shown in Figure 6, and the fact that although multifunctional, BcPMH (for which more kinetic data is available with different substrates, see Table S1 and ref 12) shows up to  $\sim 25,000$ -fold differences in  $k_{\text{cat}}/K_{\text{M}}$  values toward different substrates. These range from  $0.59 \text{ M}^{-1}\cdot\text{s}^{-1}$  for the sulfate monoester to  $1.5 \times 10^4 \text{ M}^{-1}\cdot\text{s}^{-1}$  for the phosphonate monoester. Following from this, there is also large sensitivity among alkaline phosphatases to the nature of the leaving group. For example, in the case of BcPMH, simply changing the leaving group to phenol substantially reduces the catalytic activity for all substrates by up to 350-fold (in terms of  $k_{\text{cat}}/K_{\text{M}}$ ). In the case of AP, for which linear free energy relationships do exist, these also show moderate-to-strong leaving group dependence (see, e.g., refs 22, 79, and 80). In particular, reported literature values for the AP-facilitated hydrolysis of sulfate monoesters phosphorothioates, phosphate monoesters and phosphate diesters all show steep leaving group dependence, with  $\beta_{\text{lg}}$  values in the range from  $-0.76$  to  $-0.95$ .<sup>22,79,80</sup> Therefore, although the transition states are superficially very similar, the enzyme is actually highly selective between the different substrates and leaving groups.

As these enzymes are all metalloenzymes, the catalytic metal center plays a major role in substrate positioning in all cases, guiding the ultimate orientation of the electrophile relative to the nucleophile. This is further facilitated by the involvement of a number of key residues that are strategically positioned to assist in overall substrate positioning, which for the PMHs studied here are N78, H218, and K337. These residues primarily interact with the leaving group or nonbridging oxygens of the relevant substrates and keep the electrophile in similar positions relative to the nucleophile for different substrates and electrophiles. However, due to the large binding pocket ( $\approx 10 \times 20 \text{ \AA}^2$  wide and  $15 \text{ \AA}$  deep),<sup>12</sup> there is extensive space to accommodate variations in binding conformations of spectator and leaving groups, which in turn would facilitate the accommodation of a broader range of substrates shapes. Such substrate repositioning through diversity in placement of leaving and spectator groups would therefore also play a role in determining the resulting overall catalytic efficiency and promiscuity. This is in line with our computational evidence, which highlights the importance of cooperative electrostatic interactions, brought about by active site plasticity, in accommodating a range of different substrates.

Tying in with this, if plasticity is important for facilitating promiscuity in these enzymes, one can ask whether flexibility would be reduced for catalysis by enzymes that show high specificity for their physiological substrates. A classical set of proteins where rigidity is very important in ligand binding specificity vs promiscuity are the periplasmic binding proteins, as illustrated by the structure of the cellulose-binding protein from *Thermotoga maritima*.<sup>81</sup> This protein has a bipartite active site, comprised of a solvent excluded region involved in highly specific ligand binding and is adjacent to a second and more flexible solvent-filled cavity in which semi-specific ligand binding occurs. Detailed studies of these systems as well as exploration of the role of water molecules have provided important insight into how the interplay between flexibility and rigidity allows both specificity and promiscuity to be encoded into a single binding site, moving beyond a single highly specific and fixed protein scaffold.

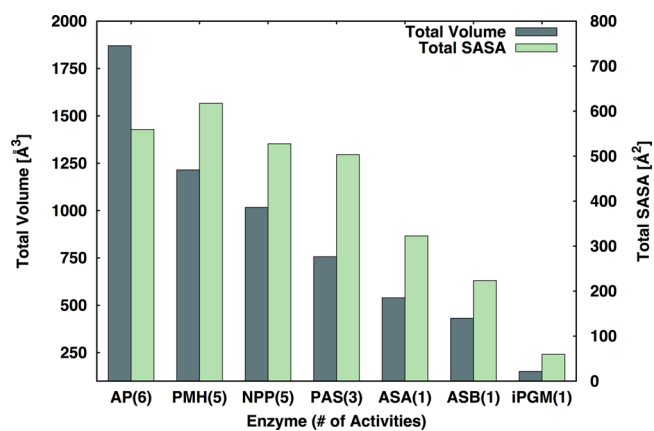
Other examples of highly specific enzymes that show comparably “rigid” active sites (in the substrate-bound conformation) are orotidine 5'-monophosphate decarboxylase<sup>82</sup> and  $\beta$ -phosphoglucomutase,<sup>83</sup> among others. These enzymes can exist in more than one conformational form and undergo ligand-gated conformational changes, engulfing their substrates upon ligand binding. However, once the substrate is bound, crucial tight binding hydrogen-bonding networks are involved in keeping the key catalytic residues in place, and, for example, truncation of various functional groups on the respective substrates can lead to tremendous reductions in catalytic activity due to the loss of key stabilizing interactions (see, for example, Richards' “substrate-in-pieces” studies of these systems that quantify the contribution of different parts of the substrate to binding and catalysis). Yet another example is a recent study of the evolution of  $\beta$ -lactamases from their promiscuous ancestral variants to their modern specific counterparts (such as TEM-1  $\beta$ -lactamase).<sup>84</sup> This work demonstrates that, within these enzymes, evolution from a generalist to a specialist enzyme is coupled to a loss of conformational flexibility.

Finally, as pointed out by a reviewer, it should be noted that the experimental work on which the present study is based<sup>8,12</sup> was performed using nonphysiological substrates, where the leaving group is the weakly basic nitrophenoxide anion (as the physiological substrate has not been identified).<sup>12</sup> The relatively small requirement for stabilizing negative charge at weakly basic anions will in turn increase the contribution of the enhanced reactivity of the active site nucleophile to the total rate accelerations outlined in Table S1, facilitating the turnover of a broader range of substrates.

## ■ IMPLICATIONS FOR EVOLUTION IN THE ALKALINE PHOSPHATASE SUPERFAMILY

As discussed in the Introduction, the phenomenon of catalytic promiscuity appears to be common among members of the AP superfamily.<sup>15</sup> Although the members of this family (which include AP, NPP, PMH, and PAS) have diverged considerably, they still share considerable similarities in active site architecture and key catalytic residues, as highlighted in Figure 1 and discussed in ref 15. In the present work, we have performed a detailed computational study of the hydrolysis of a range of substrates by two PMH, demonstrating the key role of cooperative electrostatic interactions at the PMH active site. Moreover, there is no significant conformational change in the active site alongside the reaction coordinate when comparing

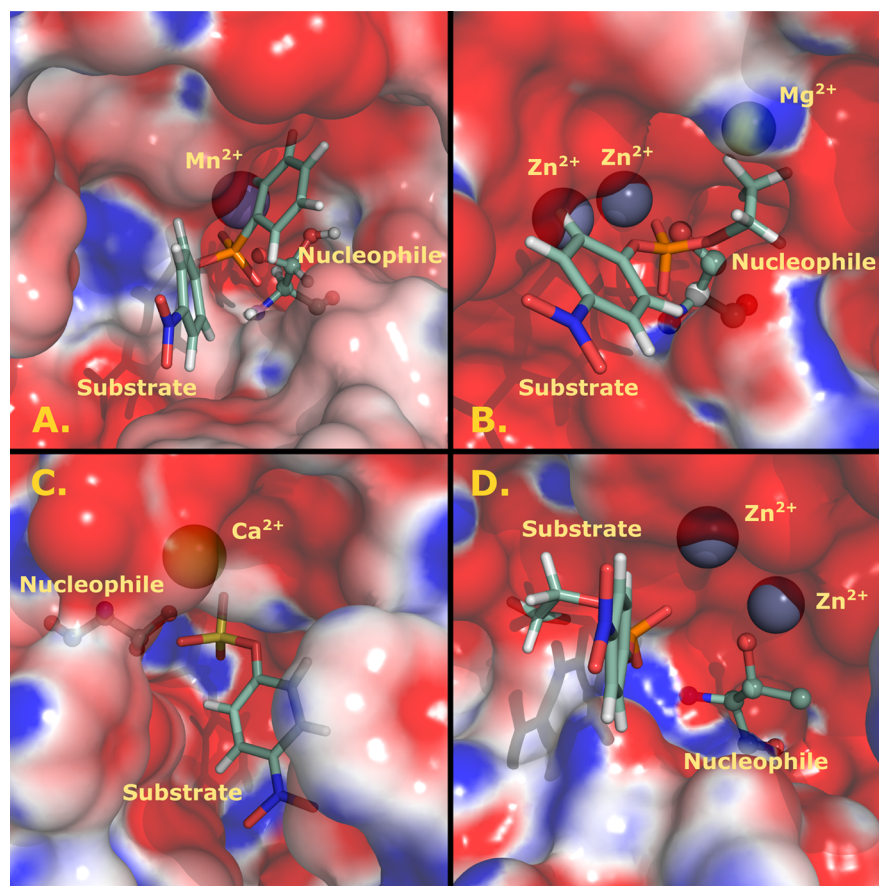
the different reactions studied. This suggests an important role for electrostatic rather than conformational<sup>86</sup> active site plasticity in facilitating the observed promiscuity in these enzymes. In order to examine this effect, we have studied the geometry of the active site cavity for different members of the AP superfamily, using the Fpocket 2 software package.<sup>87</sup> We have inspected several members of this superfamily, as well as related promiscuous phosphatases, in the search for a possible correlation between the physical properties of the different active sites and their corresponding catalytic promiscuity. The corresponding results are summarized in Figure 9 and Tables



**Figure 9.** Correlation between the number of known catalytically activities (in parentheses) and the total volume and SASA for several members of the AP superfamily.<sup>85</sup>

S7 and S8. From this data, it can be seen that when comparing different AP superfamily members, clear trends emerge with respect to the active site volumes and the number of activities that have been reported for each enzyme to date.<sup>7–10,12,21,28,71,80,88,89</sup> Specifically, according to this analysis, pockets with a larger polar surface allow the enzyme to exploit distinct residue conformations to create an optimal electrostatic environment and accommodate different transition states. In addition to this, the large volume of the different pockets would allow for the accommodation of a more diverse range of substrates, which can then be hydrolyzed through cooperative enzyme–substrate electrostatic interactions in the corresponding active sites.

Through this comparison, we find that AP, NPP, and the PMHs have the largest active site volumes and polar solvent accessible surface areas (SASAs), mostly due to the width of their active sites. Tying in with this, the arylsulfatases and other members of the superfamily have comparably smaller and narrower pockets. Interestingly, this can be directly correlated to the number of reported activities in the literature (Figure 9 and Table S7) where AP has both the largest and most accessible active site as well as the highest number of reported activities.<sup>7,10,80,88,89</sup> This is closely followed by NPP and the PMHs, with five clear activities each (not including the anomalous PTE activity of *BcPMH* for reasons outlined above). Finally, a BLAST search with *RIPMH* against known protein structures yielded the related alkaline phosphatase, PAS, as the protein with the highest sequence identity.<sup>8</sup> As can be seen from Figure 9, upon moving from the PMHs to PAS, the active site starts to reduce in volume to give a much narrower pocket, in line with the lower number of reported activities for PAS.<sup>9,28</sup> The remaining enzymes all have smaller active site volumes



**Figure 10.** Surface representation of the active sites of (A) *Burkholderia caryophylli* PMH, (B) *Escherichia coli* AP, (C) PAS, and (D) *Xanthomonas axonopodis* NPP (PDB IDs: 2W8S, 1ALK, 1HDH, and 2GSN, respectively), displaying the polar character inside the pocket. (B) and (D) show a strong presence of negatively charged residues (red) near the metal site. For both (A) and (C), there are more apolar residues present (white), although *BcPMH* still has a larger number of polar residues in its active site, mostly due to the very large size of the binding pocket.

compared to AP, NPP, and PMH, and all of them, according to Braunschweig Enzyme Database,<sup>90</sup> have so far been reported to have only one activity each. When this observation is combined with the similar mechanisms and highly polar active site residues that members of this superfamily possess (Figure 10) as well as with the experimental evidence for the existence of catalytic backups in PON1,<sup>68</sup> this strongly suggests that the cooperative electrostatic flexibility observed in PMH and related enzymes is a common feature for evolution in the AP superfamily as well as related phosphotransferases.

One limitation of this analysis, however, is that while we consider the total number of characterized activities, it neither takes into account the possibility of further as-yet uncharacterized activities in these enzymes nor the relative proficiency of these enzymes toward their promiscuous substrates. For example, even though AP has the highest number of known activities among the enzymes we examine, only two of those six activities (phosphate and phosphothioate monoester hydrolysis) are particularly proficient with  $k_{\text{cat}}/K_{\text{M}}$  values of  $3.3 \times 10^7$  and  $2.0 \times 10^4 \text{ M}^{-1}\cdot\text{s}^{-1}$ , while the other activities can have  $k_{\text{cat}}/K_{\text{M}}$  values as low as  $10^{-3} \text{ M}^{-1}\cdot\text{s}^{-1}$ .<sup>8</sup> However, clearly, a very high number of polar residues in the active site, as shown in Figure 10 and Table S7, as well as very large active sites, would allow for the presence of multiple distinct catalytic backups and a shifting electrostatic field upon substrate binding that can accommodate substrates of other shapes and charge distributions. Additionally, it is of course useful to consider not only

the total binding interactions available for transition-state stabilization but also the binding interactions that are actually needed to account for a given observed enzymatic rate acceleration, because nonspecificity will be favored whenever these total possible interactions greatly exceed the number of required interactions. For example, in the case of phosphate monoester hydrolysis, strong interactions between the enzyme and heavily charged reacting phosphate may be more than sufficient to account for the enzymatic rate acceleration. This would favor a lack of specificity for the leaving group and, perhaps, also floppiness in transition-state binding, provided there is no strict requirement of the precise placement of enzymatic side chains around the phosphate. In contrast, there may be a greater requirement for precision in the binding of the less highly charged transition state for sulfate monoester hydrolysis, which would also tie in with the experimental observation that PAS, a native sulfatase, is a more proficient phosphatase than the corresponding phosphatases in the AP superfamily are sulfatases.<sup>15</sup>

This would be supported by our observed correlation between larger active site volume/polar SASA and a great number of activities and is in sharp contrast to enzymes such as orotidine 5'-monophosphate decarboxylase, which is highly selective for the decarboxylation of orotidine monophosphate, because the enzyme makes use of every possible interaction with OMP in the stabilization of the decarboxylation state.<sup>91</sup> Interestingly, there appears to be also some sort of correlation

between tertiary structure and the number of reported activities, in that the enzymes with the highest number of characterized activities shown in Figure 9, namely AP, NPP, and PMH, are dimers and a tetramer, respectively, whereas all other enzymes are monomeric. One would assume that having a large number of polar residues in the active site would result in electrostatic strain that would have to be compensated elsewhere in the structure, which could potentially be correlated to the oligomeric states of these proteins. Overall, however, the clear correlation between increased promiscuity and a larger active site volume and SASA highlights the crucial importance of substrate charge and active site electrostatics in facilitated selectivity and evolution among these highly promiscuous enzymes.

## OVERVIEW AND CONCLUSIONS

In the present work, we have performed a detailed EVB study of both the native and several promiscuous activities of two PMH, BcPMH and R1PMH, as well as R1PMH variants with mutations in key active site residues. Our calculations can reproduce key experimental observables such as experimentally observed activation barriers for the wild-type reactions of all substrates and qualitative mechanistic predictions based on examining pH-rate profiles as well as energetic trends upon mutation of key active-site residues in R1PMH. We demonstrate that despite their broad promiscuity, both PMHs studied in this work hydrolyze all five chemically distinct substrates through a unified mechanism, binding substrates in similar positions and without the need for any significant local or global conformational changes. Additionally, we demonstrate that the apparent resilience of these enzymes to active site mutations as well as the overall promiscuity is due to compensatory electrostatic effects from different residues, allowing enough flexibility in the electrostatic environment of the active site to accommodate multiple substrates with distinct transition states and charge distributions. Finally, we provide a detailed structural and physical comparison of a range of highly promiscuous members of the AP superfamily.

These results demonstrate the strong correlation between the structural and electrostatic features of these enzyme's active sites and the corresponding variations in both substrate charge preference and the number of known promiscuous activities. This further supports our hypothesis by strongly suggesting that active site shape, size, and more critically number of polar residues available can be directly correlated the ability to accommodate increasing numbers of promiscuous activities. Our simulations and comparative analysis therefore highlight the importance of cooperative electrostatic interactions and an electrostatically flexible active site as a common feature in the evolution of promiscuous side reactions among members of the AP superfamily. In the present work we demonstrate that, in addition to the electrostatic preorganization originally suggested by Warshel in 1978,<sup>35</sup> the active site can also electrostatically reorganize to accommodate the needs of different substrates. This provides a classical example of protein flexibility allowing the reaction to occur, as these enzymes do not know in advance what substrate is going to bind. Rather, they adjust their active site environment to a given substrate after the binding step. These insights, in turn, helps us not only to understand protein evolution within a superfamily at the molecular level but also highlights a concrete feature that can be manipulated in targeted artificial enzyme design.

## ASSOCIATED CONTENT

### Supporting Information

Additional analysis performed for this work. It includes a set of tables displaying information regarding the geometry of the transition state and charges. The whole set of parameters used for all the simulations (partial charges and force field parameters), including tables with all the ionized residues, and reference values used for the energetics of every reaction studied herein to calibrate the EVB semiempirical parameters as well as derivation of values where it was possible. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b03945.

## AUTHOR INFORMATION

### Corresponding Author

\*kamerlin@icm.uu.se

### Author Contributions

‡These authors contributed equally.

### Funding

No competing financial interests have been declared.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

All calculations were performed on the High-Performance Computing Center North (HPC2N) through SNIC grant 2013/26-1. We would like to thank Profs. John Richard and Florian Hollfelder as well as Drs. Bert van Loo, Christopher Bayer, and Moshe Ben-David for valuable discussion. The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 306474. The authors would also like to acknowledge support from the Swedish Foundation for Internationalization in Higher Education and Research (STINT 2012–2097) for facilitating collaborative activity. S.C.L.K. acknowledges a personal fellowship from the Swedish Research Council (VR) No. 2010–5026, and P. B. a personal fellowship from the Sven and Lilly Lawski Foundation.

## REFERENCES

- (1) Khersonsky, O.; Tawfik, D. S. *Annu. Rev. Biochem.* **2010**, *79*, 471–505.
- (2) Jensen, R. A. *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.
- (3) O'Brien, P. J.; Herschlag, D. *Chem. Biol.* **1999**, *6*, R91–R105.
- (4) Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D. *Nature* **2006**, *440*, 1078–1082.
- (5) Roodveldt, C.; Tawfik, D. S. *Biochemistry* **2005**, *44*, 12728–12736.
- (6) Aharoni, A.; Gaidukov, L.; Yagur, S.; Toker, L.; Silman, I.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 482–487.
- (7) O'Brien, P. J.; Herschlag, D. *Biochemistry* **2001**, *40*, 5691–5699.
- (8) Jonas, S.; van Loo, B.; Hyvönen, M.; Hollfelder, F. *J. Mol. Biol.* **2008**, *384*, 120–136.
- (9) Babbie, A. C.; Bandyopadhyay, S.; Olguin, L. F.; Hollfelder, F. *Angew. Chem., Int. Ed.* **2009**, *48*, 3692–3694.
- (10) O'Brien, P. J.; Herschlag, D. *J. Am. Chem. Soc.* **1998**, *120*, 12369–12370.
- (11) Jonas, S.; Hollfelder, F. *Pure Appl. Chem.* **2009**, *81*, 731–742.
- (12) van Loo, B.; Jonas, S.; Babbie, A. C.; Benjdia, A.; Berteau, O.; Hyvönen, M.; Hollfelder, F. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2740–2745.
- (13) Luo, J.; van Loo, B.; Kamerlin, S. C. L. *FEBS Lett.* **2012**, *586*, 1622–1630.

- (14) Kim, A.; Benning, M. M.; OkLee, S.; Quinn, J.; Martin, B. M.; Holden, H. M.; Dunaway-Mariano, D. *Biochemistry* **2011**, *50*, 3481–3494.
- (15) Mohamed, M. F.; Hollfelder, F. *Biochim. Biophys. Acta* **2013**, *1834*, 417–424.
- (16) Lassila, J. K.; Zalatan, J. G.; Herschlag, D. *Annu. Rev. Biochem.* **2011**, *80*, 669–702.
- (17) Kamerlin, S. C. L.; Sharma, P. K.; Prasad, R. B.; Warshel, A. Q. *Rev. Biophys.* **2013**, *1*, 1–132.
- (18) Duarte, F.; Åqvist, J.; Williams, N. H.; Kamerlin, S. C. L. *J. Am. Chem. Soc.* **2015**, *137*, 1081–1093.
- (19) Boltes, I.; Czapinska, H.; Kahnert, A.; von Bülow, R.; Dierks, T.; Schmidt, B.; von Figura, K.; Kertesz, M. A.; Usón, I. *Structure* **2001**, *9*, 483–491.
- (20) Kim, E. E.; Wyckoff, H. W. *J. Mol. Biol.* **1991**, *218*, 449–464.
- (21) Zalatan, J. G.; Fenn, T. D.; Brunger, T. A.; Herschlag, D. *Biochemistry* **2006**, *45*, 9788–9803.
- (22) Zalatan, J. G.; Herschlag, D. *J. Am. Chem. Soc.* **2006**, *128*, 1293–1303.
- (23) Koutsioulis, D.; Lyskowski, A.; Maki, S.; Guthrie, E.; Feller, G.; Bouriotis, V.; Heikinheimo, P. *Protein Sci.* **2010**, *19*, 75–84.
- (24) Lopéz-Canut, V.; Roca, M.; Bertrán, J.; Moliner, V.; Tuñón, I. *J. Am. Chem. Soc.* **2011**, *133*, 12050–12062.
- (25) Hou, G.; Cui, Q. *J. Am. Chem. Soc.* **2012**, *134*, 229–246.
- (26) Gijssbers, R.; Ceulemans, H.; Stalmans, W.; Bollen, M. *J. Biol. Chem.* **2001**, *276*, 1361–1368.
- (27) Lopéz-Canut, V.; Roca, M.; Bertrán, J.; Moliner, V.; Tuñón, I. *J. Am. Chem. Soc.* **2010**, *132*, 6955–6963.
- (28) Olguin, L. F.; Askew, S. E.; O'Donoghue, A. C.; Hollfelder, F. *J. Am. Chem. Soc.* **2008**, *130*, 16547–16555.
- (29) Marino, T.; Russo, N.; Toscano, M. *Chem.—Eur. J.* **2013**, *19*, 2185–2192.
- (30) Hanson, S. R.; Best, M. D.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2004**, *43*, 5736–5763.
- (31) Selmer, T.; Hallmann, A.; Schmidt, B.; Sumper, M.; von Figura, K. *Eur. J. Biochem.* **1996**, *238*, 341–345.
- (32) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
- (33) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
- (34) Pauling, L. *Nature* **1948**, *161*, 707–709.
- (35) Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 5250–5254.
- (36) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.
- (37) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14115–14120.
- (38) Burschowsky, D.; van Eerde, A.; Okvist, M.; Kienhofer, A.; Kast, P.; Hilvert, D.; Kregel, U. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17516–17521.
- (39) Fried, S. D.; Bagchi, S.; Boxer, S. G. *Science* **2014**, *346*, 1510–1514.
- (40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (41) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12–21.
- (42) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Åqvist, J.; Kamerlin, S. C. L. *J. Phys. Chem. B* **2014**, *118*, 4351–4362.
- (43) Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860–2868.
- (44) Ben-David, M.; Sussman, J. L.; Maxwell, C. I.; Szeler, K.; Kamerlin, S. C. L.; Tawfik, D. S. *J. Mol. Biol.* **2015**, *427*, 1359–1374.
- (45) Sánchez-Tarín, M.; Świderek, K.; Roca, M.; Tuñón, I. *J. Phys. Chem. B* **2015**, *119*, 1899–1911.
- (46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (47) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647–3661.
- (48) Olsson, M. H. M.; Søndergard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (49) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- (50) Jorgensen, W. L.; Maxwell, D. S.; Tiado-Rives, J. J. *Am. Chem. Soc.* **1996**, *118*, 1125–11236.
- (51) Marelus, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graph. Model.* **1998**, *16*, 213–225.
- (52) *Schrödinger Release 2013–3*, MacroModel version 10.2; Schrödinger LLC: New York, 2013.
- (53) Kahn, K.; Bruice, T. C. *J. Comput. Chem.* **2002**, *23*, 977–996.
- (54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Rev. C1; Gaussian, Inc.: Wallingford, CT, 2009.
- (55) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
- (56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (57) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297–5311.
- (58) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.
- (59) Edwards, D. R.; Lohman, D. C.; Wolfenden, R. V. *J. Am. Chem. Soc.* **2012**, *134*, 525–531.
- (60) Sorensen, P. E.; Jencks, W. P. *J. Am. Chem. Soc.* **1987**, *109*, 4675–4690.
- (61) Note that there is no published experimental data for the hydrolysis of PNS and PPS by RPPMH, however, we have included these substrates in our calculations for completeness. The data plotted in this figure are presented in Table 1, and the error bars represent standard deviations over 10 independent trajectories.
- (62) Kirby, A. J.; Jencks, W. P. *J. Am. Chem. Soc.* **1965**, *87*, 3209–3216.
- (63) Bourne, N.; Williams, A. *J. Org. Chem.* **1984**, *49*, 1200–1204.
- (64) Kolmodin, K.; Åqvist, J. *FEBS Lett.* **2001**, *498*, 208–213.
- (65) Purcell, J.; Hengge, A. C. *J. Org. Chem.* **2005**, *70*, 8437–8442.
- (66) Muegge, I.; Tao, H.; Warshel, A. *Protein Eng.* **1997**, *10*, 1363–1372.
- (67) Adamczyk, A. J.; Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9827–9832.
- (68) Ben-David, M.; Elias, M.; Filippi, J.-J.; Duñach, E.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *J. Mol. Biol.* **2012**, *418*, 181–196.
- (69) Nikolic-Hughes, I.; O'Brien, P. J.; Herschlag, D. *J. Am. Chem. Soc.* **2005**, *127*, 9314–9315.
- (70) Zalatan, J. G.; Fenn, T. D.; Brunger, A. T.; Herschlag, D. *Biochemistry* **2006**, *45*, 9788–9803.
- (71) Lassila, J. K.; Herschlag, D. *Biochemistry* **2008**, *47*, 12853–12859.
- (72) Positive values correspond to destabilizing contributions and negative values to stabilizing contributions. For clarity, only the residues that contribute 1 kcal·mol<sup>-1</sup> or more to the calculated activation barrier are shown in this figure.
- (73) Baxter, N. J.; Blackburn, G. M.; Marston, J. P.; Hounslow, A. M.; Cliff, M. J.; Bermel, W.; Williams, N. H.; Hollfelder, F.; Wemmer, D. E.; Walther, J. P. *J. Am. Chem. Soc.* **2008**, *130*, 3952–3958.

- (74) Barabas, O.; Nemeth, V.; Bodor, A.; Perczel, A.; Rosta, E.; Kele, Z.; Zagyva, I.; Szabadka, Z.; Grolmusz, V. I.; Wilmanns, M.; Vertessy, B. G. *Nucleic Acids Res.* **2013**, *41*, 10542–10555.
- (75) Hou, G. H.; Cui, Q. *J. Am. Chem. Soc.* **2013**, *135*, 10457–10469.
- (76) Copley, S. *Curr. Opin. Chem. Biol.* **2003**, *7*, 265–272.
- (77) Duarte, F.; Amrein, B. A.; Kamerlin, S. C. L. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11160–11177.
- (78) Duarte, F.; Geng, T.; Marloie, G.; Al Hussain, A. O.; Williams, N. H.; Kamerlin, S. C. L. *J. Org. Chem.* **2014**, *79*, 2816–2828.
- (79) Nikolic-Hughes, I.; Rees, D.; Herschlag, D. *J. Am. Chem. Soc.* **2004**, *126*, 11814–11819.
- (80) Hollfelder, F.; Herschlag, D. *Biochemistry* **1995**, *34*, 12255–12264.
- (81) Munshi, P.; Stanley, C. B.; Ghimire-Rijal, S.; Lu, X.; Myles, D. A.; Cuneo, M. J. *BMC Struct. Biol.* **2013**, *13*, 18.
- (82) Appleby, T. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2005–2010.
- (83) Dai, J.; Finci, L.; Zhang, C.; Lahiri, S.; Zhang, G.; Peisach, E.; Allen, K. N.; Dunaway-Mariano, D. *Biochemistry* **2009**, *2009*, 1984–1995.
- (84) Zou, T. S.; Risso, V. A.; Gavira, J. A.; Sanchez-Ruiz, J. M.; Ozkan, S. B. *Mol. Biol. Evol.* **2015**, *32*, 132–143.
- (85) The enzymes chosen for this analysis were *Escherichia coli* alkaline phosphatase (AP, PDB code 1ALK), *Burkholderia caryophylli* phosphonate monoester hydrolase (PMH, PDB code 2W8S), *Xanthomonas axonopodis* nucleotide pyrophosphatase/phosphodiesterase (NPP, PDB code 2GSN), *Pseudomonas aeruginosa* arylsulfatase (PAS, PDB code 1HDH), *Homo sapiens* lysosomal arylsulfatase A (ASA, PDB code 1AUK), *Homo sapiens* lysosomal arylsulfatase B (ASB, PDB code 1FSU) and *Bacillus stearothermophilus* cofactor-independent phosphoglycerate mutase (iPGM, PDB code 1EJJ). The data plotted in this figure are presented in Table S7.
- (86) Tokuriki, N.; Tawfik, D. S. *Science* **2009**, *324*, 203–207.
- (87) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinform.* **2009**, *10*, 168.
- (88) Sun, L.; Martin, D. C.; Kantrowitz, E. R. *Biochemistry* **1999**, *38*, 2842–2848.
- (89) Zalatan, J. G.; Fenn, T. D.; Herschlag, D. *J. Mol. Biol.* **2008**, *384*, 1174–1189.
- (90) Schomburg, I.; Chang, A.; Ebeling, C.; Gremse, M.; Heldt, C.; Huhn, G.; Schomburg, D. *Nucleic Acids Res.* **2004**, *32*, D431–D433.
- (91) Spong, K.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2013**, *135*, 18343–18346.