

## Predicting Substrates by Docking High-Energy Intermediates to Enzyme Structures

Johannes C. Hermann,<sup>‡</sup> Eman Ghanem,<sup>†</sup> Yingchun Li,<sup>†</sup> Frank M. Raushel,<sup>†</sup>  
John J. Irwin,<sup>\*‡</sup> and Brian K. Shoichet<sup>\*‡</sup>

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, MC 2550, 1700 4th Street, San Francisco, California 94158-2330, and Department of Chemistry, P.O. Box 30012, Texas A&M University, College Station, Texas 77842-3012

Received August 11, 2006; E-mail: jji@cgl.ucsf.edu; shoichet@cgl.ucsf.edu

**Abstract:** With the emergence of sequences and even structures for proteins of unknown function, structure-based prediction of enzyme activity has become a pragmatic as well as an interesting question. Here we investigate a method to predict substrates for enzymes of known structure by docking high-energy intermediate forms of the potential substrates. A database of such high-energy transition-state analogues was created from the KEGG metabolites. To reduce the number of possible reactions to consider, we restricted ourselves to enzymes of the amidohydrolase superfamily. We docked each metabolite into seven different amidohydrolases in both the ground-state and the high-energy intermediate forms. Docking the high-energy intermediates improved the discrimination between decoys and substrates significantly over the corresponding standard ground-state database, both by enrichment of the true substrates and by geometric fidelity. To test this method prospectively, we attempted to predict the enantioselectivity of a set of chiral substrates for phosphotriesterase, for both wild-type and mutant forms of this enzyme. The stereoselectivity ratios of the six enzymes considered for those four substrate enantiomer pairs differed over a range of 10- to 10 000-fold and underwent 20 switches in stereoselectivities for favored enantiomers, compared to the wild type. The docking of the high-energy intermediates correctly predicted the stereoselectivities for 18 of the 20 substrate/enzyme combinations when compared to subsequent experimental synthesis and testing. The possible applications of this approach to other enzymes are considered.

### Introduction

With the sequencing of many genomes completed or underway, an emerging challenge is the functional annotation of the majority of enzymes whose activities are unknown. Several bioinformatic strategies are now widely used, including sequence analysis and gene context analysis.<sup>1,2</sup> These approaches rely on similarity to enzymes of known function. This is a sensible but sometimes unreliable strategy, since for many proteins the nearest neighbors of known activity are so dissimilar in sequence that even the same reaction mechanism cannot be assumed, while for others the operon context is uninformative. This has led to many misannotated proteins across all databases.<sup>3</sup> With the advent of the structural genomics projects, the structures of many proteins are now available well before their activities are known; even more structures are available through comparative modeling of close homologues.<sup>4</sup> For these proteins, it is a pleasant conceit to imagine that one might predict activities on the basis of structure.

One way to exploit structures for activity is with molecular docking. Docking is widely used to predict inhibitors by screening large compound databases for molecules that complement the structures of the target enzyme. Despite the method's well-known liabilities, such structure-based virtual screens have had important successes.<sup>5-11</sup> Compared to inhibitor screens, two additional problems may be anticipated when docking for a substrate. First, the quality of the docked pose is more important for substrate prediction. Second, the only complex accessible to standard docking is the ground-state Michaelis complex, and this complex is neither directly competent for turnover nor the

<sup>‡</sup> University of California, San Francisco.

<sup>†</sup> Texas A&M University.

(1) Gerlt, J. A.; Babbitt, P. C. *Genome Biol.* **2000**, *1*, REVIEWS0005.

(2) Levy, E. D.; Ouzounis, C. A.; Gilks, W. R.; Audit, B. *BMC Bioinformatics* **2005**, *6*, 302.

(3) Devos, D.; Valencia, A. *Trends Genet.* **2001**, *17*, 429-431.

(4) Pieper, U.; Eswar, N.; Davis, F. P.; Braberg, H.; Madhusudhan, M. S.; Rossi, A.; Marti-Renom, M.; Karchin, R.; Webb, B. M.; Eramian, D.; Shen, M. Y.; Kelly, L.; Melo, F.; Sali, A. *Nucleic Acids Res.* **2006**, *34*, D291-295.

(5) Specker, E.; Bottcher, J.; Heine, A.; Sotriffer, C. A.; Lilie, H.; Schoop, A.; Muller, G.; Griebenow, N.; Klebe, G. *J. Med. Chem.* **2005**, *48*, 6607-6619.

(6) Schapira, M.; Abagyan, R.; Totrov, M. *J. Med. Chem.* **2003**, *46*, 3045-3059.

(7) Rao, M. S.; Olson, A. J. *Proteins* **1999**, *34*, 173-183.

(8) Jorgensen, W. L. *Science* **2004**, *303*, 1813-1818.

(9) Li, C.; Xu, L.; Wolan, D. W.; Wilson, I. A.; Olson, A. J. *J. Med. Chem.* **2004**, *47*, 6681-6690.

(10) Schnecke, V.; Kuhn, L. A. *Perspect. Drug Discovery Des.* **2000**, *20*, 171-190.

(11) Fernandes, M. X.; Kairys, V.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1961-1970.

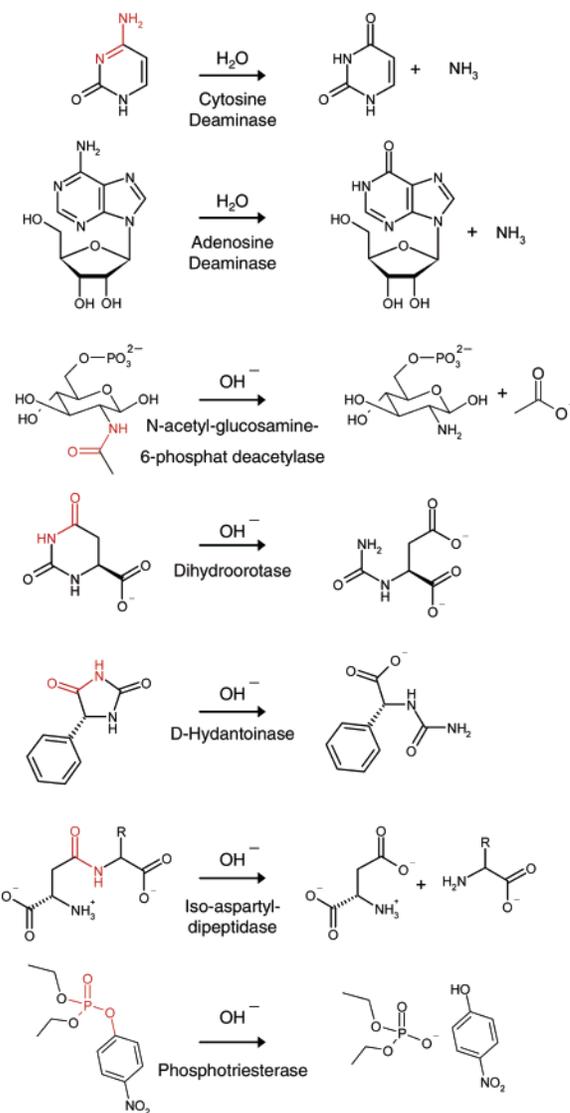
form of the substrate that the enzyme structure is pre-organized to recognize.<sup>12</sup>

To use docking for substrate prediction, one must consider the chemical transformation that the target enzymes might catalyze, a source of potential substrates that can undergo such transformations, and the form of the substrates that is preferentially recognized by the enzyme. The problem of modeling possible chemical transformations can be intimidating, since enzymes catalyze so many reactions. Similarly, the number of potential substrates is almost unbounded. Database docking typically uses ground-state chemical structures, and it is unclear that these are the most appropriate structures with which to model possible substrates. Indeed, docking ground-state substrate structures has proven problematic for activity prediction. Macchiarulo and colleagues found that, in substrate docking, the cognate ligands and enzymes had little computed specificity, leading them to propose that at least some of the observed specificity owed to nonmolecular features, such as cellular localization.<sup>13</sup> Kalyanaraman and colleagues found that standard docking alone was unsuccessful for most cases at recognizing ground-state substrate structures among a large database of decoys. Computationally more expensive rescoring did improve rankings and specificity recognition considerably, but only for the substrates that had catalytically reasonable geometries.<sup>14</sup>

If in its most general form this problem seems daunting, it is also possible to imagine simplifications that might make a docking screen for substrates pragmatic. The problem of reaction chemistry can be restricted to that of a particular family of enzymes. In this study we limit ourselves to the amidohydrolases, a functionally diverse enzyme superfamily that typically shares many mechanistic features.<sup>15,16</sup> As a source of possible substrates, one can limit sampling to those compounds likely to be encountered by the enzymes of interest. A sensible set of such potential substrates is the KEGG metabolites database, supplemented by dipeptides and several other classes of molecules, as this covers many of the molecules that amidohydrolases are known or likely to act on.<sup>16</sup> At the same time, this metabolite database provides a wide enough range of nonsubstrates as to present a challenge for docking screens. Thus, for any given enzyme, the true substrates make up less than 0.1% of the overall database—the rest of the metabolites contain functional groups that could be acted on by known amidohydrolase reactions but overall are not recognized by the particular enzymes that we target. Perhaps most importantly, we dock the metabolites not in their ground-state, stable forms but as high-energy intermediates. For instance, a tetrahedral structure was calculated for any molecule with an amine–carbon bond that was a potential substrate for conversion into a carbonyl (as in the reaction catalyzed by cytosine deaminase, Scheme 1). Similarly, any molecule with a phosphate (or phosphonate or phosphinate)–ester substructure (and thiol analogues) was represented in a trigonal-bipyramidal, pentavalent intermediate form as a potential substrate for phosphotriesterase (Scheme 1).

Here we investigate the plausibility of structure-based substrate prediction by docking high-energy intermediate forms of

**Scheme 1.** Reactions Catalyzed by Members of the Amidohydrolase Superfamily Considered in This Paper; Reaction Centers Are Colored Red



the KEGG metabolites into seven members of the amidohydrolase superfamily. We ask whether docking an activated form of these molecules is better suited to substrate prediction than the more traditional ground state, and whether these methods can prioritize likely substrates over the vast number of decoys present in the collection of known metabolites. We also turn the technique toward prospective calculation by predicting the enantioselectivities of the amidohydrolase phosphotriesterase and five of its mutant enzymes. These predictions are subsequently tested by the synthesis and testing of four potential substrates and comparison of the experimental and predicted enantiomeric preferences for phosphotriesterase and the active-site variants.

## Methods

**Overview.** To predict substrates, we analyzed the known reactions catalyzed by the amidohydrolases and selected representative reactions. On the basis of these reactions, a database of likely substrates in their ground-state forms is transformed into high-energy structures. Two independent dockable databases are thus created, one that contains the

(12) Warshel, A.; Florian, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5950–5955.

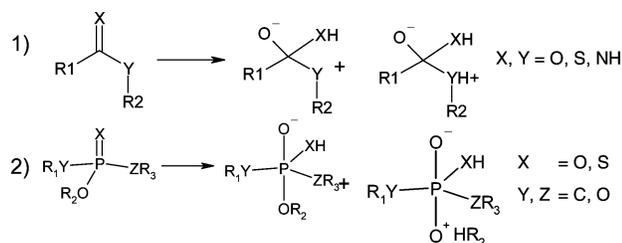
(13) Macchiarulo, A.; Nobeli, I.; Thornton, J. M. *Nat. Biotechnol.* **2004**, *22*, 1039–1045.

(14) Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Biochemistry* **2005**, *44*, 2059–2071.

(15) Holm, L.; Sander, C. *Proteins* **1997**, *28*, 72–82.

(16) Seibert, C. M.; Raushel, F. M. *Biochemistry* **2005**, *44*, 6383–6391.

**Scheme 2.** Transformation of Ground-State Structures (Left) into High-Energy Intermediate Structures Used by Docking<sup>a</sup>



<sup>a</sup> All of the high-energy forms docked, with neutral and protonated leaving groups, are represented (stereoisomers not shown). The general forms recognized by the transformation scripts are shown.

ground states and one that contains the high-energy intermediates. Both databases are then screened against the target enzymes.

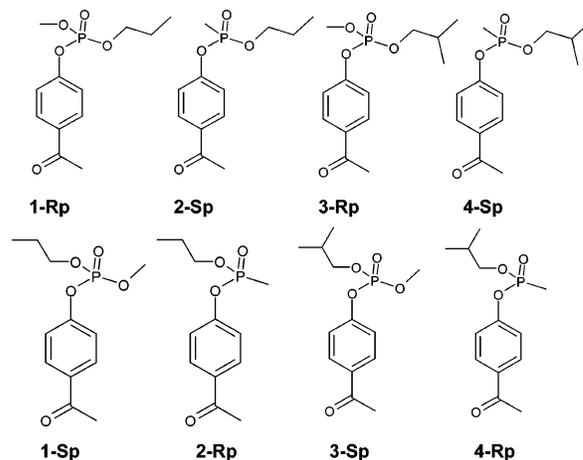
**Choice of Reactions.** To date, about 25 different reactions have been characterized among the amidohydrolases.<sup>16</sup> Almost all of the structurally characterized enzymes in this superfamily catalyze hydrolytic reactions involving a catalytic water molecule or hydroxide acting as the nucleophile (Scheme 1). A single exception is uronate isomerase (URI), which catalyzes an aldose/ketose isomerization. We restricted ourselves to the hydrolytic reactions, dividing these into three groups: hydrolysis of amide and peptide bonds, aromatic deamination reactions, and the cleavage of phosphorus esters. Besides the reactions involving heteroatoms known to occur in substrates of the amidohydrolases, we also considered all variants involving oxygen or sulfur as either the leaving group heteroatom or the carbonyl heteroatom. These variants are plausible amidohydrolase substrates, though they have not, as yet, been observed within this superfamily.

**High-Energy and Ground-State Forms of the KEGG Metabolites.** We used the KEGG metabolite database,<sup>17</sup> supplemented with dipeptides, several hydantoin, and phosphorus esters, as a source of likely substrates for the enzymes. KEGG was filtered for size and functionality. Only molecules possessing a defined electrophilic substructure, based on our chosen reactions, and fewer than 48 heavy atoms were retained. This reduced the  $\sim 13\,000$  molecules in KEGG to 3770 molecules (the version of KEGG available in June 2005 was used). We generated two independent dockable databases: ground-state forms and high-energy forms of the same molecules.

To calculate high-energy structures, all potential reactive substructures were identified. Each such substructure was separately transformed into high-energy intermediate forms using SMILES level representations of the molecules and tools from the OEChem library (OpenEye, Santa Fe, NM). These high-energy structures represent the state of the molecule after being attacked by the catalytic hydroxide, and they all contained an added hydroxyl (Scheme 2). These high-energy structures were further modified to reflect the electronics of the transition state. To do so, we made the oxygen of the added hydroxyl negative by moving its proton to the heteroatom that was double-bonded to the reactive center in the ground-state structure. For amides and esters this heteroatom was a carbonyl oxygen, whereas for cyclic amidines it was the ring nitrogen (e.g., the substrates of dihydroorotase (DHO) and cytosine deaminase (CDA), respectively, Scheme 1). This gives an electronic distribution consistent with high-level calculations of hydroxide-assisted hydrolysis, in which the attacking oxygen is negatively charged, while keeping the net charge of the intermediates correct.<sup>18,19</sup>

Multiple high-energy intermediates were generated per ground-state structure, as every reactive substructure was processed independently. Among the 3770 filtered molecules from KEGG, about 15 000 attackable substructures were identified, each of which was transformed

**Chart 1.** Potential Substrates Used To Test Docking Enantioselectivity Predictions of Phosphotriesterase and Its Mutant Enzymes



into a high-energy form. As different directions of water or hydroxide attack result in different stereoisomers for each chiral center created, the number of high-energy intermediates nearly doubled. This number expands even more dramatically for phosphorus esters, whose tetrahedral center becomes bipyramidal. In this bipyramidal geometry, the attacking hydroxide—now bound to the phosphorus atom—and the leaving group must be apical to one another. Thus, a single chiral phosphate ester with three different leaving groups can generate as many as six different intermediates. To account for potential protonation upon bond cleavage, we constructed for each leaving group in a metabolite an intermediate structure with the leaving group neutral, as it is in the ground states, and an intermediate structure in which it is protonated. An example is the amino group or the ring nitrogen in the tetrahedral intermediate of cytosine (Scheme 1). This expanded the number of high-energy intermediate structures further still. In total, there were 21 000 distinct high-energy forms of the 3770 ground-state metabolites.

The KEGG metabolites under-represent potential substrates for isoaspartyl dipeptidase (IAD), d-hydantoinase (HYD), and phosphotriesterase (PTE). When docking against these enzymes, we therefore supplemented the metabolites with dipeptides, hydantoin, and phosphorus esters, respectively. All possible combinations of L-, D-, and iso-dipeptides supplemented by all *N*-acetyl, *N*-succinyl, *N*-formyl, *N*-carbamoyl, *N*-formimino, and *N*-hydantoin amino acid derivatives were calculated, leading to 2009 peptides. Similarly, ground states and high-energy forms of 14 d-hydantoin and 5 phosphorus esters were generated. For enantioselectivity predictions for phosphotriesterase, high-energy forms of four new chiral substrates were also created (Chart 1).

**Dockable Database Preparation.** A flexibase containing multiple conformations of every molecule was used for both the high-energy and ground-state databases.<sup>20</sup> All structures were converted from isomeric SMILES to 3D structures using CORINA (Molecular Networks, Erlangen, Germany).<sup>21</sup> Flexible rings were sampled using CORINA, with a maximum of 10 ring conformations per molecule.<sup>22</sup> CORINA was also used to generate stereoisomers for stereochemically ambiguous centers. Conformations were sampled with OMEGA (Openeye, Santa Fe, NM),<sup>23</sup> using an rmsd cutoff of 0.3 Å for molecules with fewer than 21 heavy atoms, and 0.7 Å otherwise. Different protonation states of ionizable groups were created using IONIZER

(17) Kanehisa, M.; Goto, S. *Nucleic Acids Res.* **2000**, *28*, 27–30.

(18) Lopez, X.; Mujika, J. I.; Blackburn, G. M.; Karplus, M. *J. Phys. Chem. A* **2003**, *107*, 2304–2315.

(19) Hermann, J. C.; Ridder, L.; Holtje, H. D.; Mulholland, A. J. *Org. Biomol. Chem.* **2006**, *4*, 206–210.

(20) Lorber, D. M.; Shoichet, B. K. *Curr. Top. Med. Chem.* **2005**, *5*, 739–749.

(21) Sadowski, J.; Schwab, C. H.; Gasteiger, J. In *Computational Medicinal Chemistry and Drug Discovery*; Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J. P., Eds.; Dekker Inc.: New York, 2003; pp 151–212.

(22) Sadowski, J. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 53–60.

(23) Bostrom, J.; Greenwood, J. R.; Gottfries, J. J. *Mol. Graphics Modell.* **2003**, *21*, 449–462.

from the software package LIGPREP (Schrödinger, New York, NY). For groups with a  $pK_a$  value between 5 and 9, both possible protonation states were calculated; for groups outside that range, only the appropriate deprotonated or protonated form was generated. Partial atomic charges and desolvation energies were calculated using the semiempirical quantum mechanics program AMSOL.<sup>24–27</sup>

**Receptor Preparation.** Enzyme structures were prepared for docking as previously described.<sup>28</sup> Polar hydrogens were placed with SYBYL6.9 (Tripos, St. Louis, MO); AMBER charges were assigned to receptor atoms.<sup>29</sup> We had to change one aspect of our protocol when docking the high-energy versus the ground-state structures. In the former, the nucleophilic hydroxide has become part of the attacking molecule, whereas in the latter it remains as a ligating group of the catalytic metal(s). Thus, when docking the high-energy structures, the catalytic hydroxide was removed from the protein structures, but when docking the ground-state molecules, this hydroxide was retained. As in our previous docking study of metallo-enzymes, we found it important to redistribute the formal 2+ charge on each metal atom to the ligating residues.<sup>28</sup> For bi-metallo centers (DHO, HYD, NAGA, IAD, PTE), charges of 1.3 and 1.4, or 1.4 and 1.5, for hydroxide- and non-hydroxide-containing active sites were assigned to the  $\alpha$  and  $\beta$  zinc ions, respectively (the different charges on the two metals are attributable to their different ligation patterns). The remaining charge of 0.5–0.7 per metal ion was distributed over the ligating residues, similar to a previously reported protocol.<sup>28</sup> For mono-metallo enzymes (CDA, ADA), the hydroxide was replaced by a water molecule for docking ground-state structures; it was removed for docking high-energy forms. In these two enzymes, the base, which abstracts a proton from the water molecule upon nucleophilic attack, was represented in its protonated form (for CDA His246; for ADA His238).<sup>16</sup> A charge of 1.4 or 1.3 was assigned to the metal, depending on whether the water molecule is present or not. The remaining net formal charge on the metal was distributed to the ligating residues.

Docking grids for van der Waals interaction energies and excluded volume were calculated using CHEMGRID and DISTMAP, respectively.<sup>30</sup> DELPHI was used to calculate electrostatic potential grids using an internal dielectric of 2 and an external dielectric of 78.<sup>31</sup> The dielectric of a defined region in the active site potentially occupied by substrate atoms was also set to 2 to account for the effect of substrate binding, as described.<sup>32</sup> A set of manually curated spheres generated by SPHGEN was used to orient molecules in the binding site.<sup>32</sup>

For the prediction of the enantioselectivity of phosphotriesterase (PTE), five mutant enzyme structures were modeled. The mutant enzymes differed from the wild type by only one to three active-site residues. These residues have, compared to the wild-type enzyme, either a slightly larger or smaller side chain projecting into the active site. Since a crystal structure of a related phosphotriesterase mutant differs from the wild-type protein by only 0.47 Å RMSD in the  $\alpha$ -carbons and had no major backbone movements, the mutant enzymes were modeled by simple modifications of the side chains of the wild-type structure.<sup>33</sup> We held the overall structure constant and retained the

coordinates of the backbone and common atoms of wild type and the newly created residues.

**Docking Procedure.** All docking runs were performed using DOCK3.5.54.<sup>30</sup> Initial ligand orientations were sampled using receptor and ligand bin sizes of 0.5 Å and a ligand and receptor overlap of 0.4 Å. The distance tolerance for matching a receptor and a ligand sphere was set to 1.5 Å. Using these rather aggressive matching parameters, up to one million initial poses per molecule were generated, for each of which multiple conformations were scored. The best scoring pose was rigid-body minimized<sup>34</sup> and scored for electrostatic and van der Waals interactions. Configurations were penalized for desolvation using the precalculated AMSOL ligand desolvation energies, weighted for the degree of burial of each substrate atom by the protein (B. Shoichet, unpublished).

All docked poses were filtered for consistency with our model of catalysis; molecules docked in a catalytically nonproductive configuration were rejected. This was performed by simply insisting that reactive substructures be within 3.5 Å of the catalytic nucleophile. Only the best scoring representation of each molecule was considered in the final hit list. Thus, for both the ground-state and high-energy KEGG databases, the final ranked hit list had a maximum of 3770 molecules.

**Synthesis of Racemic Substrates.** Compounds 1–4 (Chart 1) were synthesized in racemic mixtures and tested following the same procedure as described in detail in the accompanying paper (Nowlan et al.<sup>46</sup>).

4-Acetylphenyl methyl propanyl phosphate (**1**) was prepared from methyl dichlorophosphate and propanol. <sup>1</sup>H NMR (ppm): 7.96 (2H, d,  $J = 8.55$  Hz), 7.30 (2H, d,  $J = 8.55$  Hz), 4.17–4.08 (2H, m), 3.87 (3H, d,  $J_{H-P} = 11.40$  Hz), 2.58 (3H, s), 1.76–1.67 (2H, m), 0.95 (3H, t, 7.49 Hz). <sup>31</sup>P NMR (ppm): –4.87. Mass spectrometry (M + H): calculated, 257.09; found, 257.10.

4-Acetylphenyl propanyl methylphosphonate (**2**) was prepared from methylphosphonic dichloride and propanol. <sup>1</sup>H NMR (ppm): 7.98 (2H, d,  $J = 8.54$  Hz), 7.32 (2H, d,  $J = 8.54$  Hz), 4.20–4.00 (2H, m), 2.60 (3H, s), 1.77–1.65 (2H, m), 1.68 (3H, d,  $J_{H-P} = 17.63$  Hz), 0.96 (3H, t, 7.45 Hz). <sup>31</sup>P NMR (ppm): 28.54. Mass spectrometry (M + H): calculated, 273.09; found, 273.09.

4-Acetylphenyl isobutanyl methyl phosphate (**3**) was prepared from methyl dichlorophosphate and isobutanol. <sup>1</sup>H NMR (ppm): 7.96 (2H, d,  $J = 8.55$  Hz), 7.30 (2H, d,  $J = 8.55$  Hz), 3.96–3.91 (2H, m), 3.87 (3H, d,  $J_{H-P} = 11.40$  Hz), 2.58 (3H, s), 2.02–1.93 (1H, m), 0.94 (3H, t, 6.28 Hz). <sup>31</sup>P NMR (ppm): –4.90. Mass spectrometry (M + H): calculated, 287.17; found, 287.17.

4-Acetylphenyl isobutanyl methylphosphonate (**4**) was prepared from methylphosphonic dichloride and isobutanol. <sup>1</sup>H NMR (ppm): 7.94 (2H, d,  $J = 8.54$  Hz), 7.28 (2H, d,  $J = 8.54$  Hz), 3.94–3.78 (2H, m), 2.57 (3H, s), 2.01–1.83 (1H, m), 1.66 (3H, d,  $J_{H-P} = 17.73$  Hz), 0.91 (6H, d,  $J = 76.77$  Hz). <sup>31</sup>P NMR (ppm): 28.52. Mass spectrometry (M + H): calculated, 271.11; found, 271.11.

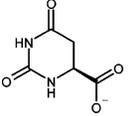
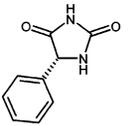
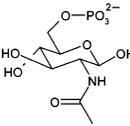
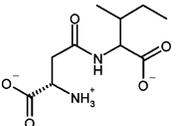
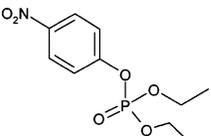
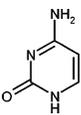
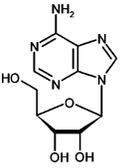
## Results

**Overview.** In retrospective docking calculations, both the high-energy structures and the ground states were docked into the seven control amidohydrolases for which the substrate preferences were known (Table 1). The ability to identify the known substrates was evaluated on the basis of their ranking in the docking hit list. In these rankings, the scores of the annotated substrates are compared to those of the 3770 other molecules in the database, which are considered nonsubstrate decoys. Since only well-oriented potential substrates pass the post docking filters and appear in the hit list, the quality of docked poses is taken into account when comparing the results.

- (24) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385–16398.
- (25) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820–1831.
- (26) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins* **1999**, *34*, 4–16.
- (27) Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. *J. Mol. Biol.* **2006**, *357*, 1449–1470.
- (28) Irwin, J. J.; Raushel, F. M.; Shoichet, B. K. *Biochemistry* **2005**, *44*, 12316–12328.
- (29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- (30) Meng, E. C.; Shoichet, B.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505–524.
- (31) Gilson, M. K.; Honig, B. H. *Nature* **1987**, *330*, 84–86.
- (32) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (33) Hill, C. M.; Li, W. S.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. *J. Am. Chem. Soc.* **2003**, *125*, 8990–8991.

- (34) Gschwend, D. A.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 123–132.

**Table 1.** Members of the Amidohydrolase Superfamily Studied Here

enzyme	PDB ID	sample substrate
dihydroorotase (DHO)	1J79 <sup>39</sup>	dihydroorotate 
D-hydantoinase (HYD)	1GKP <sup>40</sup>	5-phenyl-hydantoin 
<i>N</i> -acetyl-glucosamine-6-phosphate-deacetylase (NAGA)	1UN7 <sup>41</sup>	<i>N</i> -acetyl-glucosamine-6-phosphate 
iso-aspartyl-dipeptidase (IAD)	1ONX <sup>42</sup>	L-iso-aspartyl-isoleucine 
phosphotriesterase (PTE)	1HZY <sup>43</sup>	paraoxon 
cytosine deaminase (CDA)	1K6W <sup>44</sup>	cytosine 
adenosine deaminase (ADA)	1A4M <sup>45</sup>	adenosine 

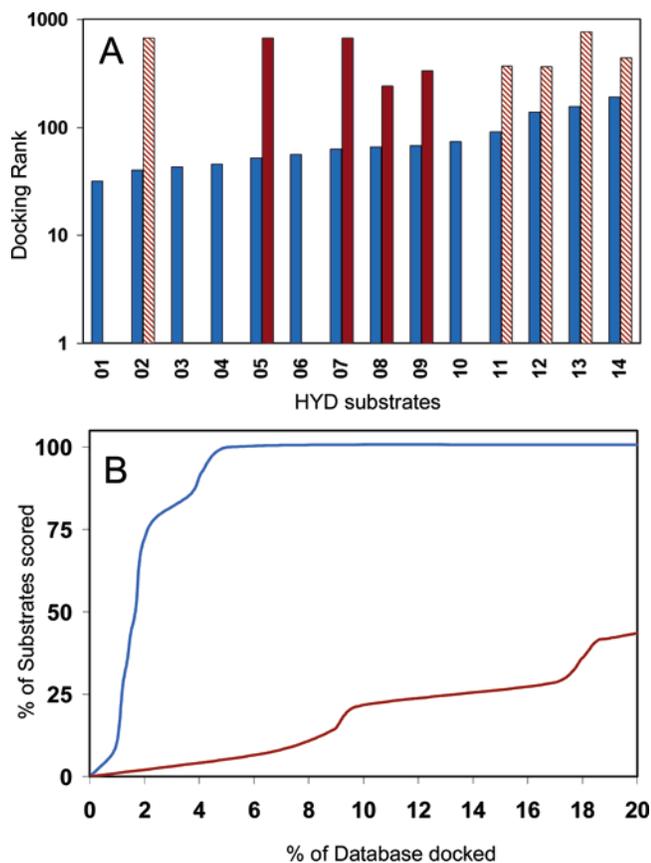
We also made prospective predictions of enantioselectivities of mutants of phosphotriesterase for four new substrates, which were tested experimentally afterward.

**Retrospective Docking.** By both ranking and geometry, docking the excited states outperformed docking the ground states in five of the seven amidohydrolases. For two out of the seven targets, dihydroorotase (DHO) and adenosine deaminase (ADA), both databases performed extraordinarily well and the differences were negligible. For most enzymes, however, including D-hydantoinase (HYD), phosphotriesterase (PTE), cytosine deaminase (CDA), *N*-acetyl-glucosamine-6-phosphate-deacetylase (NAGA), and iso-aspartyl-dipeptidase (IAD), the differences between the ground- and excited-state database were unmistakable. For these enzymes, many more high-energy intermediate structures adopted catalytically productive geometries compared with the analogous ground states, and these high-energy intermediate forms typically ranked much better compared to their respective decoys. In all seven enzymes, the high-energy forms of the known substrates ranked in the top 100 molecules out of 3770, and often they ranked among the top 10 or 20 molecules. The advantage of the high-energy database is most striking when docking to apo structures (HYD, PTE, and CDA). Here the results of the ground states are not just worse than those for the high-energy forms, but they fail

to distinguish the substrates from the substrate decoys. Below we consider each enzyme in more detail. The non-specialist may wish to skip to the next section, Prospective Docking.

For D-hydantoinase, L-iso-aspartyl dipeptidase, and phosphotriesterase, there were 14, 37, and 11 known substrates in the database, respectively, enough to calculate enrichment factors for the docking results. For D-hydantoinase, all of the 14 high-energy substrates ranked among the top 5% of the database (top 200 molecules), giving an enrichment factor of 20 at this point (Figure 1B; Supporting Information Chart S1). Conversely, none of the ground-state forms ranked among the top 5%. Only 4 of the 14 ground-state substrates docked in a catalytically productive pose, with none ranking higher than 242nd out of the entire database (Figure 1

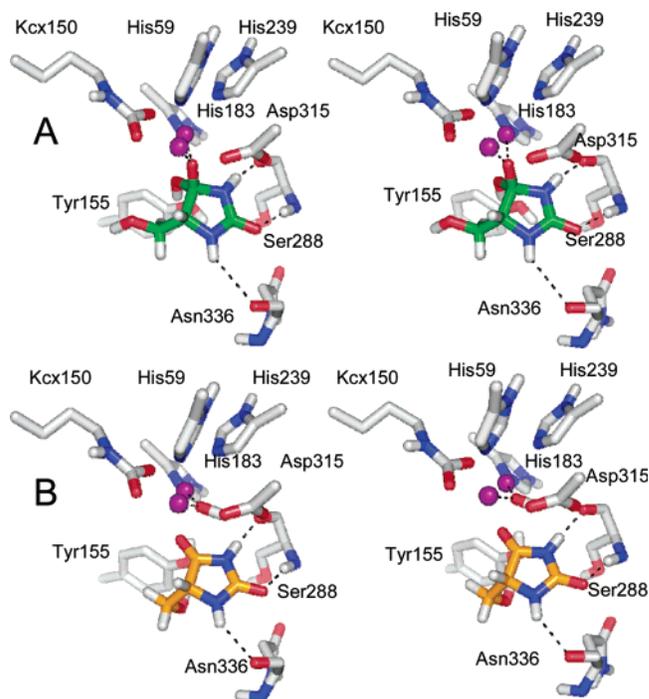
A). The tetrahedral high-energy intermediates adopt docked geometries that placed the oxyanion—the former attacking hydroxide oxygen—in a position where it ligated the two zinc centers (Figure 2A). The distances between the oxyanion and the zinc atoms are typically about 2.1 Å. The ground-state structures are oriented such that the oxygen of the reactive carbonyl group coordinates the  $\beta$ -zinc ion and the carbon is presented to the attacking hydroxide at a distance of about 2.8 Å (Figure 2B). For both types of substrate structures, the ring heteroatoms of the substrate hydrogen bond with the backbone



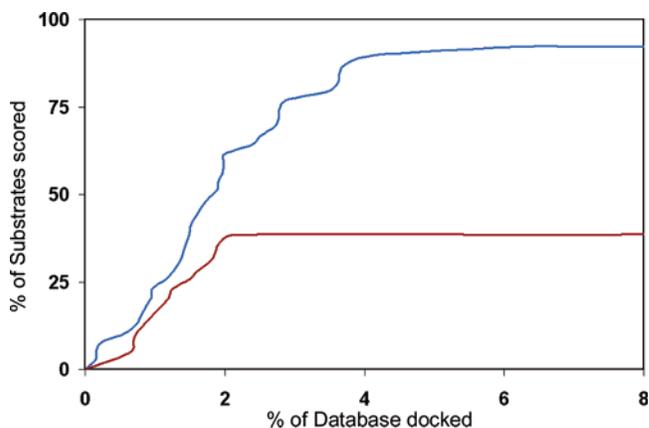
**Figure 1.** Docking results for D-hydantoinase. (A) Docking ranks of substrates (Supporting Information Chart S1) in their ground-state forms (brown columns) are compared to those of the same substrates in their high-energy forms (blue columns). Solid columns give the ranks for substrates docked in a catalytically competent pose, and striped columns indicate nonproductive geometries. (B) Enrichment plot for high-energy (blue curve) and ground-state (brown curve) databases.

atoms of Ser288 and Asn336. The higher (better) rankings of the high-energy structures versus the ground-state structures, relative to the decoys in each respective database, appears to arise from the closer approach to these backbone atoms allowed by the tetrahedral geometries in the hydantoin substrates. For instance, for 5-hydroxyethyl-hydantoin, the distance between N1 of the substrate and the backbone oxygen of Ser288 shrinks from 2.75 Å in the ground-state complex to 2.6 Å in the high-energy intermediate complex (Figure 2). A key point is that the higher ranking of the true substrates in the high-energy intermediate screens comes not directly from their anionic nor their tetrahedral forms, which both the true substrates and decoys possess, but in the better interactions made by the substrate specificity groups that the high-energy intermediate geometries allow.

For the dipeptidase, the ground-state database gives satisfactory results, but the high-energy intermediate database performs much better. In the ground-state database docking, 38.5% of the L-iso-Asp dipeptides rank among the first 2.1% of the docked database, corresponding to an enrichment factor of 18.3; all further dipeptides after this point adopt catalytically incompetent poses (Figure 3). In contrast, 92% of the high-energy L-iso-Asp dipeptides are docked in a functional orientation with 61% of them ranked among the top 2% of the database, corresponding to an enrichment factor of 30.5. Overall, 33 of the 37 L-iso-Asp dipeptide substrates rank better in their high-energy forms



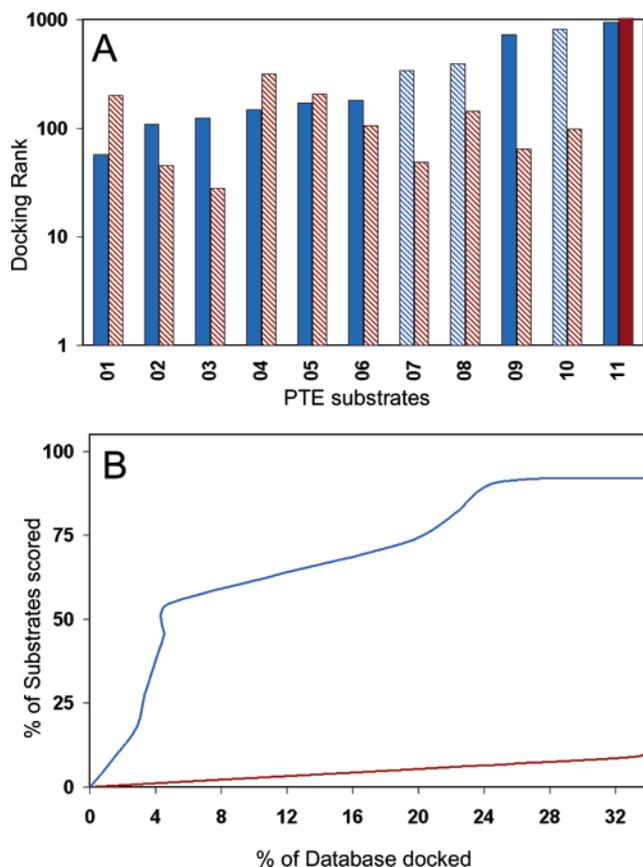
**Figure 2.** Stereoviews of 5-hydroxyethyl-hydantoin docked in its high-energy (A) and its ground-state (B) forms into D-hydantoinase. Both poses are catalytically competent. Zinc ions are shown as purple spheres, oxygen atoms are colored red, enzyme carbons in gray, ligand carbons in green (high-energy form) or orange (ground-state form), hydrogens in white, and nitrogens in blue. Hydrogen bonds are drawn as black dashed lines. The fourth ligating histidine (His61) is undisplayed. This figure and Figure 6 were rendered using PyMOL (South San Francisco, CA).



**Figure 3.** Docking results for iso-aspartyl-dipeptidase, shown as enrichment plots for the high-energy (blue curve) and ground-state (brown curve) databases.

than in the corresponding ground states (data not shown). Intriguingly, D-iso-Asp and D/L-( $\alpha$ )-Asp dipeptides also rank well; L-( $\alpha$ )-Asp dipeptides have recently been shown to be substrates for L-iso-aspartyl dipeptidase.<sup>35</sup> The high-energy intermediates are docked with the oxyanion between the two zinc ions, similar to the substrates in D-hydantoinase (Figure 2). In the docked pose, the two carboxylates and the  $\alpha$ -ammonium group of the dipeptides form apparently favorable hydrogen bonds with key active-site residues (Gly75, Glu77, Thr106, Arg169, Arg233, and Ser289). For the aspartyl moiety, these geometries correspond closely to the aspartate cocrystallized

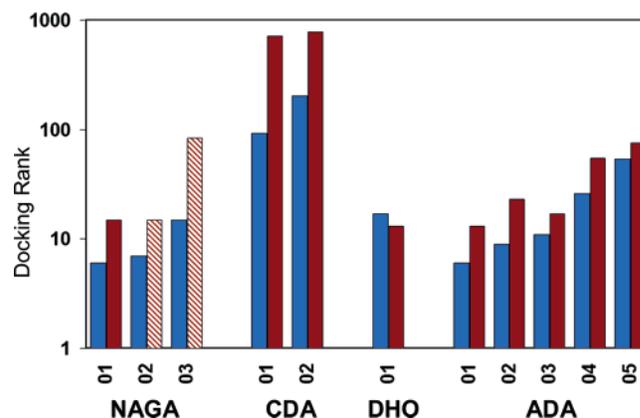
(35) Marti-Arbona, R.; Fresquet, V.; Thoden, J. B.; Davis, M. L.; Holden, H. M.; Raushel, F. M. *Biochemistry* 2005, 44, 7115–7124.



**Figure 4.** Docking results for phosphotriesterase. (A) Docking ranks of substrates (Supporting Information Chart S2) in their ground-state forms (brown columns) are compared to those of the same substrates in their high-energy forms (blue columns). Solid columns give the ranks for substrates docked in a catalytically competent pose, and striped columns indicate nonproductive geometries. (B) Enrichment plot for high-energy (blue curve) and ground-state (brown curve) databases.

in the active site of the crystal structure with a RMSD of typically about 0.6 Å. For the ground states, many nonfunctional docked poses are observed. Here, the difference may be competition among oxyanionic groups. Without a strong recognition site to orient the oxyanion toward the zinc ions, many competing nonfunctional but electrostatically reasonable poses exist, and the functional orientation for the ground states is not always the best-scoring docked pose.

For phosphotriesterase, the substrates (Supporting Information Chart S2) ranked well in both high-energy intermediate and ground-state database screens, with several from each database ranking among the top 100 (Figure 4A). However, only one of the ground-state structures adopted a catalytically competent geometry, whereas eight of the high-energy substrate structures did so. Accounting for geometry in both cases, therefore, led to much better enrichment factors for the high-energy database docking than the ground-state docking; 50% of the annotated substrates from the former are found among the top 5% of the docking-ranked database, which gives an enrichment factor of 10 at this point (Figure 4B). Conversely, the only ground-state substrate to adopt a catalytically competent pose ranked poorly and is not found within the top 33% of the database. The better docking orientations and rankings of the high-energy intermediates are explained by the combination of an anionic form and the disposition of substrate functionality in a trigonal-bipyra-

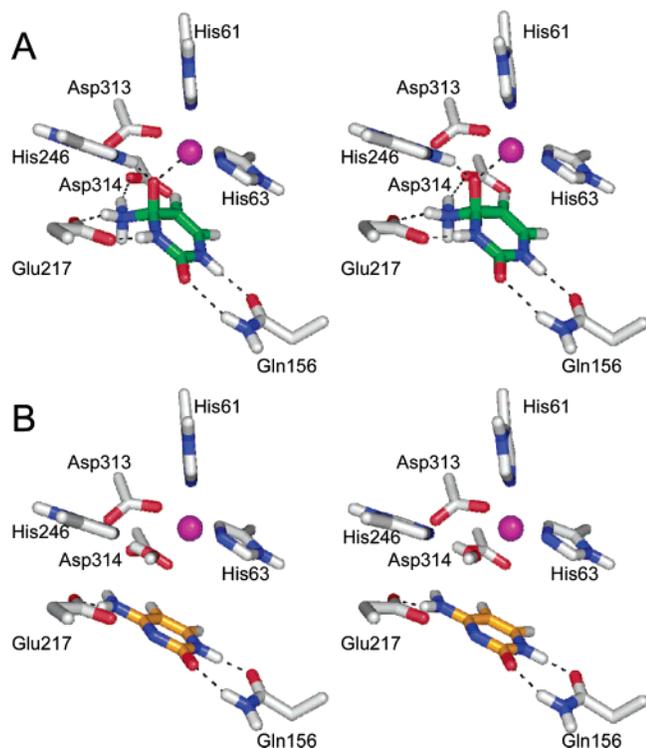


**Figure 5.** Docking results for *N*-acetyl-glucosamine-6-phosphate-deacetylase (NAGA), cytosine deaminase (CDA), dihydroorotase (DHO), and adenosine deaminase (ADA). Docking ranks of substrates (Supporting Information Chart S3) in their ground-state forms (brown columns) are compared to those of the same substrates in their high-energy forms (blue columns). Solid columns give the ranks for substrates docked in a catalytically competent pose, and striped columns indicate nonproductive geometries.

midal, as opposed to a tetrahedral, geometry. The specificity of phosphotriesterase is mainly defined by apolar interactions between the substrate and the protein, and the shape of the active sites goes a long way toward selecting the substrate (see accompanying paper by Nowlan et al.<sup>46</sup>). The high-energy substrates dock so as to place the oxyanion between the two zinc ions, with each apolar side chain of the phosphorus fitting snugly into one of the side-chain pockets and the appropriate group apical to the oxyanion in the “leaving pocket”. The ground states, on the other hand, find several alternative, less favorable and non-catalytically competent orientations. For all but one of the best substrate orientations, either the pose does not place the electrophilic center remotely near the attacking hydroxide, or the pocket for the leaving group is occupied by a phosphorus side chain that is not hydrolyzed.

For both cytosine deaminase and NAGA, there were too few substrates to justify enrichment curves. Instead, we directly compared the two substrates of cytosine deaminase and the three substrates of NAGA as ranked by docking either the high-energy intermediates or the ground-state databases (Figure 5, Supporting Information Chart S3). For both enzymes, substrates ranked substantially better when treated as part of a high-energy intermediate database than when docked as part of a ground-state database. The three substrates for NAGA ranked within the top 15 molecules of the docking-ranked high-energy intermediate database, corresponding to 0.4% of the entire database. All were docked in a catalytically competent pose. For the ground-state database, only one out of the three substrates was docked in a productive pose, ranking 15th out of 3770. The two substrates for cytosine deaminase are docked in a catalytically competent geometry in both database screens, but they rank much better as high-energy intermediates, at 92nd and 204th, than in the ground-state forms, where they rank 711th and 778th, only within the top 20.5% and too low to be distinguished from the decoy molecules in the database.

For both enzymes, the improved performance of the high-energy forms relates to their charge distribution and structure. In NAGA, the tetrahedral high-energy structures adopt poses similar to those of the other bimetallic enzymes, with the



**Figure 6.** Stereoviews of cytosine docked in its high-energy (A) and its ground-state (B) forms into cytosine deaminase. Both poses are catalytically competent. Oxygen atoms are colored red, enzyme carbons in gray, ligand carbons in green (high-energy form) or orange (ground-state form), hydrogens in white, and nitrogens in blue. The purple sphere represents the catalytic iron ion. The third ligating histidine (His 214) is undisplayed.

oxyanion coordinated between the two zinc ions. The phosphate or sulfate side chains of the glucosamine substrates (Supporting Information Chart S3) ion-pair with His233 and Arg234. Conversely, only one of the ground states is docked in a functional orientation, with its amide functional group oriented toward the attacking hydroxide. Two of the ground-state substrates are placed in a nonfunctional pose in which the neutral amide functional group, which is transformed into a negatively charged tetrahedral structure in the high-energy forms, is pointing toward the solvent and the negatively charged side chain interacts with a zinc ion rather than with His233 and Arg234. This is a case of an internal functional group acting as a decoy zinc ligand, causing the substrate to adopt a catalytically incompetent geometry. It is interesting to note that, among the top 15 compounds of the high-energy hit list, 8 are sugar derivatives, suggesting that the structural pattern recognized by NAGA is being captured by the docking calculation when the high-energy intermediate database is used. The two high-energy substrates for cytosine are docked into the monovalent active site of cytosine deaminase so that the negatively charged oxyanion is placed between His246 and the iron atom (Figure 6A). The nitrogen attached to the electrophilic carbon is protonated and ion-pairs with Glu217. The leaving ammonium group interacts with two anionic residues (Glu217 and Asp314), with N–O(carboxylate) distances of 3 and 4 Å, respectively. The carbonyl oxygen of the substrate and the adjacent nitrogen proton hydrogen-bond to the side chain of Gln156. The complementarity of the ground states to the active site is less favorable (Figure 6B). The nitrogen adjacent to the electrophilic carbon is not protonated and cannot interact with Glu217. The

leaving amino group is neutral and cannot ion-pair with Glu217 and Asp314. Consequently, the relative ranks of the ground states are lower than those of the high-energy intermediate structures.

The two cases where the high-energy intermediates and the ground states docked equally well were two targets for which substrate docking was highly successful using either of the two databases. In dihydroorotase, both the ground and high-energy states of dihydroorotate rank highly, with the ground state ranked 13th and the high-energy form ranked 17th out of the 3770 molecules in the databases, essentially all of which are decoys (Figure 5, Supporting Information Chart S3). Both the ground-state and high-energy forms dock in orientations closely resembling that adopted by a crystallized dihydroorotate molecule (rmsd 0.6 Å). In docking to adenosine deaminase, all four substrates in the high-energy database ranked within the 26 top-scoring compounds and within the 55 top-scoring compounds in the ground-state database hit list (Figure 5, Supporting Information Chart S3). Whereas each substrate ranks better in the high-energy screen than in the ground-state screen, both forms did so well as to be effectively indistinguishable. Both enzymes (dihydroorotase and adenosine deaminase) were crystallized as holo structures, which have accommodated either the ground-state substrate (dihydroorotase), i.e. the identical molecule that is docked in the ground-state database, or a transition-state analogue inhibitor (adenosine deaminase). Such accommodation in the holo structures contributed to the quality of the docking results.

**Prospective Docking.** In addition to the retrospective calculations, where we knew the substrates in advance, we also wanted to test the method prospectively, predicting substrate enantioselectivity. We turned to the enzyme phosphotriesterase, for which we had previously calculated enantioselectivities for known substrates and mutant enzymes retrospectively (see accompanying paper by Nowlan et al.<sup>46</sup>). The mutant phosphotriesterases contain simple point substitutions of up to three residues, all of which project into the active site, primarily affecting the size of specificity pockets that accommodate nonreactive side chains for the phosphorus core of the substrate. Ligand binding and enantioselectivity are thought to be mostly affected by changes in steric fit in these mutant enzymes. For instance, in the mutant G60A, the size of the smaller of the specificity pockets is further reduced, increasing its enantioselectivity versus the wild-type enzyme.<sup>36</sup> We docked Sp- and Rp-stereoisomers of four potential new substrates into wild-type phosphotriesterase and five mutants (Chart 1). On the basis of the differences in the docking scores of the stereoisomers, the stereoselectivity for each of the six enzymes was predicted.

The four compounds (i.e., eight stereoisomers) were subsequently synthesized and tested as substrates, and the stereoselectivity of the six enzymes was determined (Table 2). The wild type preferentially hydrolyzes one of the two enantiomers for three of the four substrates, with enantiomeric preferences ranging from 170-fold (for the Sp- over the Rp-enantiomer of compound **3**) to no preference for either enantiomer for compound **2**. For compound **4** the wild-type enzyme preferentially hydrolyzed the Rp-enantiomer, whereas for compounds **1** and **3** the Sp-enantiomer was preferred. For the G60A mutant

(36) Chen-Goodspeed, M.; Sogorb, M. A.; Wu, F.; Raushel, F. M. *Biochemistry* **2001**, *40*, 1332–1339.

**Table 2.** Enantioselectivity Ratios for Wild-Type and Mutant Phosphotriesterase

compound	WT	G60A	H257Y	I106G	I106G	I106A
			L303T	H257Y	F132G	F132A
<b>1</b> (Sp) <sup>a</sup>	9	960	-6	-14	-38	-24
<b>2</b> (Rp) <sup>a</sup>	1	170	-73	-25	-65	-200
<b>3</b> (Sp) <sup>a</sup>	170	4.1 × 10 <sup>4</sup>	-14	-67	-49	-18
<b>4</b> (Rp) <sup>a</sup>	10	1.3 × 10 <sup>4</sup>	-50	-67	-200	-370

<sup>a</sup> The absolute stereochemistry of the enantiomer preferred by the WT and G60A mutant forms of PTE is indicated. For the other four mutant enzymes the opposite enantiomer is preferred, indicated by negative ratios.

**Table 3.** Stereoselectivities Predicted by Molecular Docking

	WT	G60A	H257Y	I106G	I106G	I106A
			L303T	H257Y	F132G	F132A
<b>1</b> Δ dock scores <sup>a</sup>	-6.99	-8.49	1.28	17.19	11.01	8.13
-RT ln(k <sub>Sp</sub> /k <sub>Rp</sub> ) <sup>b</sup>	-1.28	-4.00	1.04	1.54	2.12	1.85
<b>2</b> Δ dock scores <sup>a</sup>	<b>18.72</b>	19.34	-4.28	-59.21 <sup>c</sup>	-57.82 <sup>c</sup>	nd <sup>d</sup>
-RT ln(k <sub>Sp</sub> /k <sub>Rp</sub> ) <sup>b</sup>	0.00	2.99	-2.50	-1.87	-2.43	-3.08
<b>3</b> Δ dock scores <sup>a</sup>	-2.00	-4.39	0.07	15.20	12.34	5.13
-RT ln(k <sub>Sp</sub> /k <sub>Rp</sub> ) <sup>b</sup>	-2.99	-6.18	1.54	2.45	2.26	1.68
<b>4</b> Δ dock scores <sup>a</sup>	15.65	16.07	-0.54	nd <sup>d</sup>	nd <sup>d</sup>	<b>3.00</b>
-RT ln(k <sub>Sp</sub> /k <sub>Rp</sub> ) <sup>b</sup>	1.34	5.51	-2.28	-2.45	-3.08	-3.44

<sup>a</sup> The difference in docking score between the Sp- and the Rp-enantiomers (Δ dock scores = Sp-score - Rp-score, in kcal mol<sup>-1</sup>). A positive value indicates that the Sp-enantiomer is preferred in that particular mutant, whereas a negative value indicates that the Rp-enantiomer is preferred. Values are shown in bold when the docking prediction is inconsistent with the experimental results. <sup>b</sup> The difference in experimental rates of hydrolysis, given as RT ln(k<sub>Sp</sub>/k<sub>Rp</sub>), in kcal mol<sup>-1</sup> (Table 1). A positive value indicates that the Sp-enantiomer is preferred in that particular mutant, whereas a negative value indicates that the Rp-enantiomer is preferred. <sup>c</sup> Only one enantiomer of a pair was docked in a catalytically productive pose. <sup>d</sup> Not determined.

the preference of the wild-type enzyme was enhanced for every compound, whereas all other mutants had an inverted preference compared to the wild type and G60A. For instance, the Rp-enantiomer of compound **4** is hydrolyzed 10 times faster than the Sp-enantiomer by the wild-type phosphotriesterase. The G60A mutant hydrolyzes this enantiomer 1300-fold better than the Sp-enantiomer. Conversely, mutant I106A/F132A/H257W hydrolyzes the Sp-enantiomer of **4** 370-fold faster than the Rp-enantiomer, inverting the wild type and G60A preferences. The relative hydrolytic rates for the different stereoisomers of **4** for these two mutants thus differ by 4.8 × 10<sup>6</sup>.

Overall, 21 docking predictions of enantioselectivity were made for the 24 possible mutant/substrate pairs, including the wild-type enzyme. For three combinations, no predictions were made because none of the docked poses was catalytically competent. The substrates were subsequently synthesized and the enantioselectivities determined by experiment. Docking predictions were qualitatively consistent with the experimental observations in 19 of the 21 cases and wrong in two of them (Table 3). For the wild-type enzyme, the correct enantiomer was predicted as the preferred geometry for three of the four compounds; for compound **2** the Rp-enantiomer was strongly preferred by docking, even though no significant difference was observed experimentally. For the G60A mutant enzyme, the same preferences, including the preference of compound **2** for the Rp-enantiomer, were observed as predicted. Moreover, in G60A the experimental enantioselectivities increased relative

to the wild type, and this was mirrored in the increased differential docking scores between the two enantiomers. For the four other mutant enzymes, an inversion in enantioselectivity was observed for each of the four compounds relative to the wild type. The relative docking scores captured these inversions reliably, at least qualitatively. Out of the 16 inversions measured, 15 were predicted correctly and only that for compound **4** in mutant I106A/F132A/H257W was predicted incorrectly by docking.

A caveat to these results is that, whereas the docking predictions were qualitatively consistent with the experimental results, quantitatively there were large differences. In the wild-type and the G60A enzymes, for instance, the magnitude of the difference in docking energies typically exceeded that of the observed enantioselectivities. Reasons why one might hope for qualitative but discount quantitative correspondence between docking energies and experimental rate differences are considered below.

## Discussion

Although structure-based prediction of enzyme substrates is subject to several possible pitfalls, many may be avoided by pragmatic simplifications. By restricting ourselves to a single enzyme superfamily, the amidohydrolases, the possible enzyme-catalyzed reactions were narrowed from essentially infinite to around 10. By only modeling metabolites and related molecules, the possible number of substrates was reduced from the googol levels quoted for chemical space to about 4000. The largest technical innovation was the development of a method to represent and dock high-energy rather than ground-state forms of the database of potential substrates.

For substrate prediction, docking high-energy states rather than ground states seems intuitive and led to striking improvements in substrate recall over a ground-state database. The correct substrates not only rank higher when docked as high-energy intermediates than they do when docked as ground states, but they also are much more likely to adopt catalytically competent poses. Of the 72 substrates docked into the seven amidohydrolases, 42 were fit in nonproductive configurations in their ground-state forms. Conversely, only six substrates fit nonproductively in their high-energy forms. Part of this improvement undoubtedly comes from the oxyanion present in the high-energy intermediate database, which interacts with the catalytic metals of the enzymes. This is only part of the story, however, since these oxyanions are present in all of the molecules in the high-energy intermediate database, and the true substrates must outscore these decoy molecules to rank well. The distinguishing feature among the high-energy intermediates is the ability of the true substrates, in their tetrahedral or bipyramidal forms, to complement the specificity pockets of the enzyme while at the same time ligating the metal centers. In the nonsubstrate molecules the correct specificity groups are not present, or are arranged incorrectly, and so these decoys rank lower. In the ground-state substrates, where the correct functional groups are present, the oxyanion is missing and the geometry is not optimal to fit the specificity groups in their cognate pockets.

The enhanced recall of the high-energy versus the ground-state structures was unusually pronounced when docking against apo conformations of the target enzymes. This was the case

with D-hydantoinase, phosphotriesterase, and cytosine deaminase. Whereas these apo structures were more challenging targets for both types of docking molecules, they were particularly difficult for the ground states. For instance, the worst target for the high-energy intermediates was cytosine deaminase. Here the substrate, cytosine, was ranked within the top 2.5% of the database. Although worse than the performance in the holo structures, such a ranking might still identify cytosine as a plausible substrate, if it was unknown. In contrast, in the ground-state docking, cytosine and methylcytosine are not ranked within the top 20% of the database molecules. For phosphotriesterase and D-hydantoinase the difference was sharper still, as most of the ground-state docked poses were not even catalytically competent. When enzymes of unknown function are crystallized in their apo conformations, as may be the typical case, high-energy intermediates should be the first choice to probe enzyme function.

Of course, too much can be read into the apparent success of the high-energy intermediate docking, and several problems deserve mention. Although docking these excited states correlated well with experiment qualitatively, there was little quantitative correlation between, for instance, the magnitude of predicted and experimental enantioselectivity ratios nor the magnitude of predicted and experimental rate enhancements. Quantitative prediction of affinities for even inhibitors is beyond docking, and when bond-breaking reactions are involved, as they are in enzyme catalysis, the ice on which we skate becomes thin. We can only hope to model such reactions quantitatively using a much higher level of theory, such as combined quantum mechanics and molecular mechanics methods (QM/MM).<sup>37,38</sup> Using high-energy states as a proxy for reactivity works only because of the improved enzyme–substrate complementarity at this point along most reaction coordinates, and because of

our relatively good understanding of what such high-energy intermediates are likely to look like for amidohydrolase reactions. For other families of enzymes, our ability to limit substrates and reactions and anticipate high-energy structures may be more limited. Even with the amidohydrolases, we were concerned that our database of 3770 metabolites was too limited, and as the method is turned to predicting genuinely new substrates for targets of unknown function, the database of possible substrates and reactions will have to expand.

Notwithstanding these caveats, a compelling result to emerge from this study was the ability to prospectively predict enantioselectivity of new substrates for phosphotriesterase. Enough predictions were made and found to correspond with experiment (essentially 19 out of 21) as to make the chance of fortuitous correspondence remote. Nor was such prediction wholly trivial, as the enantioselectivity of the mutants changed sign from the preferences of the wild type in many enzyme/substrate pairs, and these inversions of the preference were qualitatively captured by docking the high-energy intermediates, as were all four cases where the enantioselectivity of the wild-type enzyme was enhanced by a mutant enzyme. Admittedly, our cause was favored by the importance of steric complementarity in the phosphotriesterase specificity pockets (see accompanying paper by Nowlan et al.<sup>46</sup>), and one can easily imagine cases where more subtle changes in pocket shape, dynamics, or polar complementarity would make this task more challenging. Nevertheless, these structure-based enantioselectivity predictions have few precedents and are genuinely encouraging. Application of this structure-based approach for substrate prediction to enzymes of unknown function thus merits further study. To aid such efforts, we have made our database of high-energy intermediates publicly available (<http://www.hei.docking.org>).

**Acknowledgment.** This work was supported by NIH grants GM71790 (PI J. A. Gerlt), GM71896 (to B.K.S. and J.J.I.), and GM33893 (to F.M.R.). J.C.H. thanks the “Deutschen Akademie der Naturforscher Leopoldina” for a fellowship under contract no. BMBF-LPD 9901/8-115. We thank R. Brenk for many helpful discussions and A. Graves for reading the manuscript.

**Supporting Information Available:** Exact structures of all substrates for the seven amidohydrolases used in this paper for docking; coordinate files for docked substrates, which should be looked at using the PDB files defined in Table 1. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA065860F

- (37) Claeysens, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. *Chem. Commun. (Cambridge)* **2005**, 5068–5070.
- (38) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (39) Thoden, J. B.; Phillips, G. N., Jr.; Neal, T. M.; Raushel, F. M.; Holden, H. M. *Biochemistry* **2001**, *40*, 6989–6997.
- (40) Abendroth, J.; Niefind, K.; Schomburg, D. *J. Mol. Biol.* **2002**, *320*, 143–156.
- (41) Vincent, F.; Yates, D.; Garman, E.; Davies, G. J.; Brannigan, J. A. *J. Biol. Chem.* **2004**, *279*, 2809–2816.
- (42) Thoden, J. B.; Marti-Arbona, R.; Raushel, F. M.; Holden, H. M. *Biochemistry* **2003**, *42*, 4874–4882.
- (43) Benning, M. M.; Shim, H.; Raushel, F. M.; Holden, H. M. *Biochemistry* **2001**, *40*, 2712–2722.
- (44) Ireton, G. C.; McDermott, G.; Black, M. E.; Stoddard, B. L. *J. Mol. Biol.* **2002**, *315*, 687–697.
- (45) Wang, Z.; Quijcho, F. A. *Biochemistry* **1998**, *37*, 8314–8324.
- (46) Nowlan, C.; Li, Y.; Hermann, J. C.; Evans, T.; Carpenter, J.; Ghanem, E.; Shoichet, B. K.; Raushel, F. M. *J. Am. Chem. Soc.* **2006**, *128*, 15892–15902.