

## Experimental and Computational Mutagenesis To Investigate the Positioning of a General Base within an Enzyme Active Site

Jason P. Schwans,<sup>†</sup> Philip Hanoian,<sup>‡</sup> Benjamin J. Lengerich,<sup>‡</sup> Fanny Sundén,<sup>†</sup> Ana Gonzalez,<sup>||</sup> Yingssu Tsai,<sup>||,⊥</sup> Sharon Hammes-Schiffer,<sup>\*,§</sup> and Daniel Herschlag<sup>\*,†</sup>

<sup>†</sup>Department of Biochemistry, Stanford University, B400 Beckman Center, 279 Campus Drive, Stanford, California 94305, United States

<sup>‡</sup>Department of Chemistry, Pennsylvania State University, 104 Chemistry Building, University Park, Pennsylvania 16802, United States

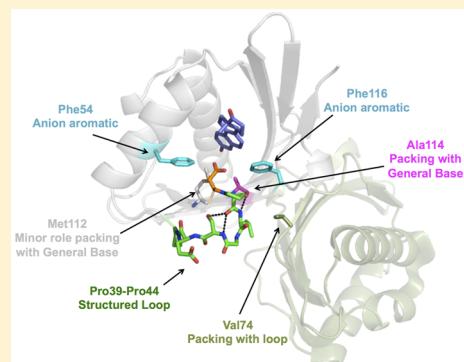
<sup>§</sup>Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, United States

<sup>||</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States

<sup>⊥</sup>Department of Chemistry, Stanford University, 333 Campus Drive, Stanford, California 94305, United States

### Supporting Information

**ABSTRACT:** The positioning of catalytic groups within proteins plays an important role in enzyme catalysis, and here we investigate the positioning of the general base in the enzyme ketosteroid isomerase (KSI). The oxygen atoms of Asp38, the general base in KSI, were previously shown to be involved in anion–aromatic interactions with two neighboring Phe residues. Here we ask whether those interactions are sufficient, within the overall protein architecture, to position Asp38 for catalysis or whether the side chains that pack against Asp38 and/or the residues of the structured loop that is capped by Asp38 are necessary to achieve optimal positioning for catalysis. To test positioning, we mutated each of the aforementioned residues, alone and in combinations, in a background with the native Asp general base and in a D38E mutant background, as Glu at position 38 was previously shown to be mispositioned for general base catalysis. These double-mutant cycles reveal positioning effects as large as 10<sup>3</sup>-fold, indicating that structural features in addition to the overall protein architecture and the Phe residues neighboring the carboxylate oxygen atoms play roles in positioning. X-ray crystallography and molecular dynamics simulations suggest that the functional effects arise from both restricting dynamic fluctuations and disfavoring potential mispositioned states. Whereas it may have been anticipated that multiple interactions would be necessary for optimal general base positioning, the energetic contributions from positioning and the nonadditive nature of these interactions are not revealed by structural inspection and require functional dissection. Recognizing the extent, type, and energetic interconnectivity of interactions that contribute to positioning catalytic groups has implications for enzyme evolution and may help reveal the nature and extent of interactions required to design enzymes that rival those found in biology.



A central goal of biology is to understand the factors responsible for the enormous rate enhancements provided by enzymes. A distinguishing feature of enzymes relative to small-molecule catalysts is the use of noncovalent interactions to position substrates and functional groups relative to one another, and many X-ray structures of enzymes reveal apparent networks of interactions. However, understanding how these structural features position active site residues, how important different features are in this positioning, and how they operate together requires extensive functional tests.

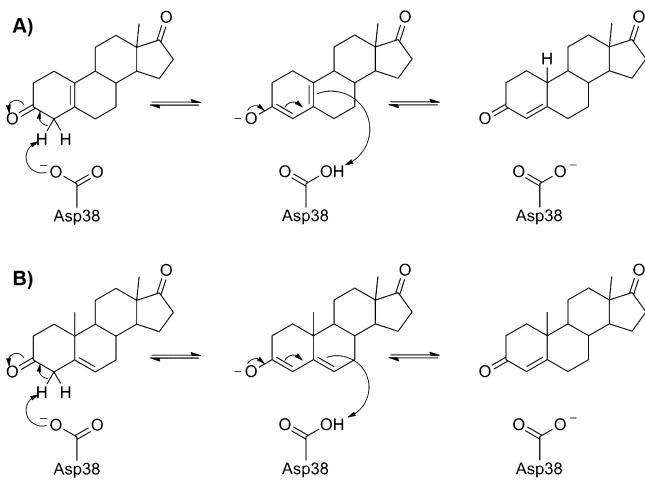
Three catalytic strategies have been identified for the enzyme ketosteroid isomerase (KSI): stabilization of the incipient oxyanion via an oxyanion hole, facilitated proton abstraction via a general base (Figure 1), and localization and positioning of the substrate via binding interactions.<sup>1–4</sup> In particular, the nature

of the hydrogen bonding interactions in the oxyanion hole has been extensively investigated using experimental<sup>5–12</sup> and theoretical<sup>13–21</sup> approaches. Recently, the oxyanion hole has been fully ablated, yet 10<sup>9</sup>-fold of the overall 10<sup>12</sup>-fold catalysis remains, indicating that other features—those noted above or hitherto unidentified catalytic sources—are responsible for the remaining billion-fold catalysis.<sup>22</sup> We previously showed that anion–aromatic interactions with Phe54 and Phe116 help position the general base carboxylate via favorable interactions between the carboxylate and the partial positive charges situated on the hydrogen atoms on the aromatic rings.<sup>23</sup> Here we

**Received:** December 16, 2013

**Revised:** February 26, 2014

**Published:** March 5, 2014



**Figure 1.** Steroid isomerization reaction of (A) 5(10)-estrene-3,17-dione [5(10)-EST] and (B) 5-androstene-3,17-dione (5-AND) catalyzed by the general base Asp38 in ketosteroid isomerase in *Comamonas testosteroni*. The isomerization process involves two sequential proton transfer reactions: transfer of a proton from C4 of the steroid to Asp38, followed by transfer of a proton from Asp38 to C6 of the steroid in 5-AND and C10 of the steroid in 5(10)-EST.

investigate the roles of more distal active site interactions to better understand how the general base is positioned.

Prior studies have suggested that residues not directly participating in the chemical transformation can play important roles in enzymatic catalysis. For example, the activity of a catalytic antibody was increased by 10<sup>2</sup>-fold through mutation of only non-active site residues, an effect that was attributed to repositioning active site residues for optimal geometries.<sup>24</sup> Mutation of loop residues ~10 Å from the active site and mutation of residues directly within the trypsin binding pocket were required to change the substrate specificity of trypsin to that of chymotrypsin.<sup>25</sup> In each study, the residues identified were not obvious *a priori* from visual inspection of the enzyme's structure. More generally, the overall folded structure of an enzyme is necessary to establish the active site architecture and to position the residues within it.<sup>26–31</sup>

Motivated by these and other prior results and the current challenge to understand the positioning of catalytic groups, we probed nearby residues that might contribute to the positioning and thus the function of the catalytic base Asp38 in KSI. Our results highlight the extended nature of catalytic sites and the need to understand effects extending beyond readily identifiable catalytic residues. The insights provided by this and related studies provide crucial information toward the development of predictive models to describe the roles of distal residues and active site features in catalysis.

## METHODS

All buffers were prepared with reagent grade materials or better. All reagents were of the highest purity commercially available (≥97%), except for the substrate 5(10)-estrene-3,17-dione [5(10)-EST] (Steraloids, Inc.). Solutions of the steroid substrate slowly converted to 4-estrene-3,17-dione (product) as determined by thin layer chromatography. The concentrations of substrate stock solutions were determined by measuring the absorbance change at 248 nm for the acetate-catalyzed conversion of substrate to product and using the molar absorptivity of the product (14800 M<sup>-1</sup> cm<sup>-1</sup>).<sup>8</sup> Independent

measurements with ≥93% pure substrate and ≥99% pure substrate agreed within experimental error, indicating ≤7% product in the substrate stock did not affect kinetics measurements. All substrate stocks contained ≤7% product contamination.

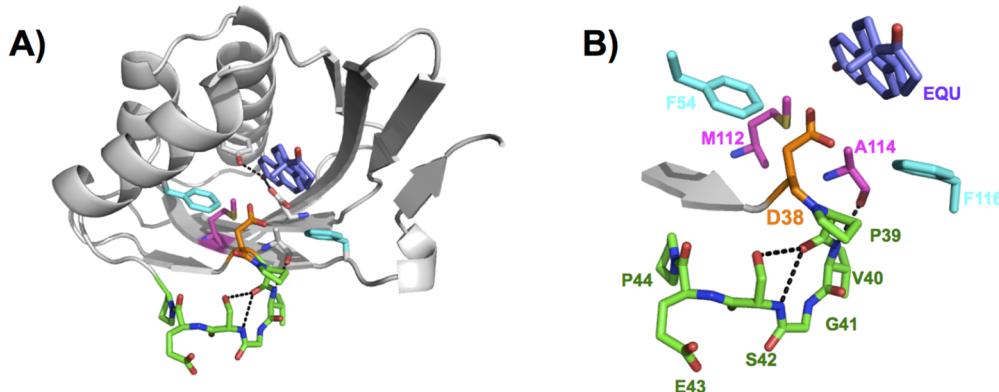
**KSI Mutagenesis, Expression, and Purification.** Quik-Change (Stratagene) site-directed mutagenesis was used to introduce the mutations into KSI genes from *C. testosteroni* encoded on pKK22-3 plasmids or pET21c plasmids. Mutations were confirmed by sequencing mini-prep DNA from DH5α cells. Proteins were expressed and purified as previously described.<sup>9</sup> The final purity was >99% as estimated on the basis of a Coomassie-stained sodium dodecyl sulfate–polyacrylamide gel electrophoresis gel. The protein concentration was determined using the calculated molar extinction coefficient in 6 M guanidinium hydrochloride according to the method of Gill et al.<sup>32</sup>

**Kinetics Experiments.** Reactions with 5(10)-EST were conducted at 25 °C in 10 mM buffer with 2% DMSO (v/v) added as a cosolvent for substrate solubility, 1 mM sodium EDTA, and 2 mM DTT and were monitored continuously at 248 nm in a PerkinElmer Lambda 25 spectrophotometer. A constant ionic strength of 100 mM was maintained in all samples using NaCl. The following buffers were used: sodium citrate, pH 4.3–5.9; sodium phosphate, pH 6.1–8.0; sodium 3-(*N*-morpholino)-propanesulfonate (MOPS), pH 6.3–8.3. The molar absorptivity for the product of 14800 M<sup>-1</sup> cm<sup>-1</sup> was previously determined using commercially available 4-estrene-3,17-dione.<sup>8</sup> The values of  $k_{cat}$ ,  $K_M$ , and  $k_{cat}/K_M$  were determined by fitting the initial rates as a function of substrate concentration (typically eight concentrations that varied from 2 to 600 μM) to the Michaelis–Menten equation. At least three determinations at differing enzyme concentrations (varying at least 4-fold overall) were averaged.

**X-ray Crystallography.** Single-crystal diffraction data were collected at SSRL beamline BL9-1 using a wavelength of 0.98 Å.<sup>33</sup> The reflections were indexed and integrated with *XDS*,<sup>34</sup> the intensities were scaled, merged, and converted to amplitudes with *SCALA* and *TRUNCATE*.<sup>35</sup> Molecular replacement was performed with *AMoRe*<sup>36</sup> using the 3CPO structure, and refinement was performed with *REFMAC5*.<sup>37,38</sup> Manual model building was conducted with *COOT*.<sup>39</sup>

**Molecular Dynamics Simulations.** Classical MD simulations of wild-type (WT) and mutant KSIs were initiated from relevant crystal structures. For WT KSI and the A114G, S42G, and V40G/S42G mutants, chains C and D of the crystal structure of KSI D38N with bound equilenin [Protein Data Bank (PDB) entry 1QJG]<sup>40</sup> were used as an initial structure. For the D38E and D38E/A114G mutants, chains A and B of unliganded D38E KSI (PDB entry 4L7K) were used.<sup>23</sup> For the P39G and D38E/P39G mutants, the unliganded structure of KSI P39A (PDB entry 3MHE) determined herein was used. For the P39G/V40G/S42G mutants, chains A and B of KSI P39G/V40G/S42G with bound equilenin (PDB entry 3OV4) were used. For D38E/P39G/V40G/S42G, the crystal structure of unliganded D38E/P39G/V40G/S42G (PDB entry 3NM2) determined herein was used.

KSI was simulated as a dimer with bound 17-oxo-andro-3,5-dien-3-olate in all cases. This ligand corresponds to the intermediate in the steroid isomerization reaction of 5-androstene-3,17-dione (5-AND), while the kinetics experiments were performed for the isomerization reaction of the 5(10)-EST substrate (Figure 1). Changes in rates due to mutation have been



**Figure 2.** General base, Asp38, positioned in the KSI active site. (A) Superposition of the WT unbound structure (PDB entry 8CHO) and the equilenin-bound D38N structure (PDB entry 1QJG), where only equilenin (EQU) is shown from the equilenin-bound structure. The Asp general base (orange) is located at the end of a  $\beta$ -strand (gray) and at the beginning of a loop (green), and it is situated near Phe54 and Phe116 (cyan) and Met112 and Ala114 (magenta). The bound equilenin ligand (blue-violet) is situated at the location of the bound substrate during the isomerization reaction. (B) Detailed view of the general base and surrounding residues with labels. Colors are as in panel A.

shown to be qualitatively consistent between these two substrates.<sup>41</sup> Furthermore, prior simulations using the intermediates of the isomerization reactions of 5-AND and 5(10)-EST were found to be qualitatively similar.<sup>42</sup>

The crystal structures were modified as needed to obtain a complete dimeric structure of the protein. For simulations initiated from the P39A crystal structure (PDB entry 3NM2), the complete chain in the asymmetric unit (chain B) was aligned by root-mean-square deviation (rmsd) minimization of  $C_{\alpha}$  atoms with respect to the 1QJG crystal structure to obtain a dimer. Similarly, for D38E/P39G/V40G/S42G, the single monomeric unit in the structure (PDB entry 3NM2) was aligned by rmsd minimization with respect to 1QJG to obtain a dimer. Note that the initial conformation of the catalytic base, residue 38, in the P39G/V40G/S42G and D38E/P39G/V40G/S42G mutants differs significantly from that of the WT and is directed more toward the solvent. This difference persists throughout the simulations and is addressed in Results and Discussion. In the 3NM2 and 3OV4 crystal structures, residues Pro39Gly, Val40Gly, Gly41, and Ser42Gly were not present and were added using Profix,<sup>43</sup> with peptide bonds in the *trans* conformation, which is consistent with that observed in the D38E/P39G/V40G/S42G structure (PDB entry 3NM2). In cases where Ala125 was absent (PDB entries 4L7K and 3OV4), this C-terminal residue was also added using Profix.

The structures were further modified to incorporate the ligand, add protons as needed, and solvate the system. In cases where a bound equilenin ligand was present (PDB entries 1QJG and 3OV4), an energy-minimized structure of the steroid intermediate (17-oxo-andro-3,5-dien-3-olate) was aligned with equilenin by minimizing the rmsd of the C3, O3, C4, C17, and O17 atoms in the steroid and the corresponding atoms in equilenin. For cases where equilenin was absent (PDB entries 4L7K and 3NM2), the steroid intermediate was aligned by minimizing the rmsd of the  $C_{\alpha}$  atoms for each monomer with respect to the equilenin-bound 1QJG crystal structure. Aspartic acid residues 38 and 99 were considered to be protonated, and the steroid intermediate was considered to be deprotonated. Histidines 6 and 100 were protonated at the  $\epsilon$ -position, and histidine 122 was doubly protonated, based on protonation states determined using  $H^{+}$ <sup>44</sup> and inferable hydrogen bonds observed in the 1QJG crystal structure. Each system was immersed in a

truncated octahedral box with 12900 explicit water molecules and four sodium counterions to ensure a neutral system.

The AMBER99SB force field<sup>45,46</sup> was employed for the protein, and the rigid TIP3P water model<sup>47</sup> was used for the water molecules. Charges for the intermediate state of the steroid were determined using the restrained electrostatic potential (RESP) method<sup>48</sup> and are the same as those reported previously.<sup>42</sup> Missing steroid force field parameters were obtained from the Generalized Amber Force Field (GAFF).<sup>49</sup> A 10 Å real space cutoff was used for nonbonded interactions, and long-range electrostatics were treated with the particle mesh Ewald method.<sup>50,51</sup> Bonds involving hydrogen were constrained to their equilibrium bond lengths using LINCS.<sup>52</sup>

A multistep equilibration procedure was performed prior to data collection for each system. First, the positions of the solvent and ions were optimized using the steepest descent energy minimization algorithm; subsequently, the solvent and ions were equilibrated at constant NVT for 500 ps at 300 K using the Nosé-Hoover thermostat.<sup>53,54</sup> Then the positions of the solvent and ions were optimized again by energy minimization, followed by optimization of the full system. The system was then annealed from 50 to 300 K, heating in increments of 50 K over 10 ps and maintaining the temperature during each interval for 50 ps at constant NPT at a pressure of 1 atm using the Nosé-Hoover thermostat and Parrinello-Rahman barostat.<sup>55</sup> The system was then simulated at constant NPT for 2 ns. A configuration consistent with the equilibrated average volume from this simulation was extracted and further equilibrated at constant NVT for 2 ns. These simulations were extended for additional production for 20 ns, and configurations were saved every 1 ps. A time step of 1 fs was used in all simulations. All simulations were performed using Gromacs.<sup>56</sup> Two independent trajectories were performed for each dimeric system, resulting in data for four active sites that were averaged to obtain the data presented in the main text; data for individual trajectories are presented in the Supporting Information.

## RESULTS AND DISCUSSION

The general base of KSI, Asp38, is situated at the end of a  $\beta$ -sheet and the beginning of a loop (Figure 2A), with its side chain proximal to those of Phe54, Met112, Ala114, and Phe116 (Figure 2B). Prior studies have suggested the presence of favorable anion–aromatic interactions between the carboxylate oxygens of

Asp38 and the hydrogen atoms on the aromatic rings of Phe54 and Phe116.<sup>23</sup> A structured loop connecting two stands of an antiparallel  $\beta$ -sheet via a  $\beta$ -turn<sup>57</sup> begins with the general base Asp38 and contains Pro39, Val40, Gly41, Ser42, Glu43, and Pro44, with a *cis* peptide linkage between Asp38 and Pro39 (Figure 2B). Our very simple expectation was that the structure of this loop and packing interactions with the nearby hydrophobic side chains would help to orient Asp38 to act as a general base. However, the underlying energetic contributions to catalysis of the structured loop and packing interactions cannot be ascertained through structural inspection alone; in particular, the structure does not indicate which residues are needed for optimal general base catalysis, which have larger or smaller contributions, or whether the entire loop or portions of the loop function as cooperative units.

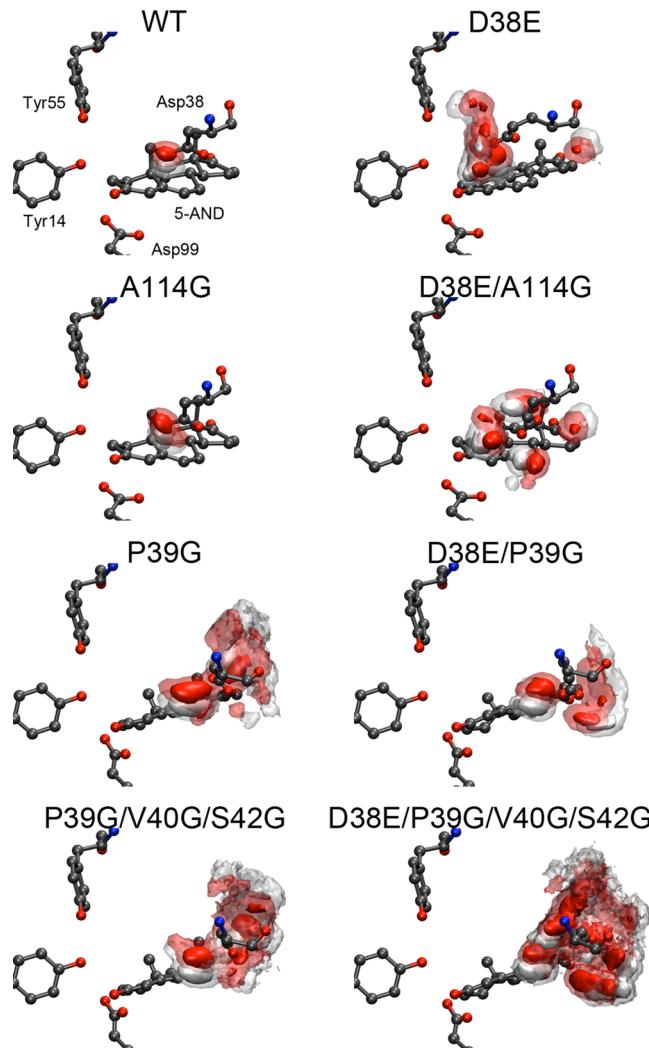
Prior kinetic studies showed that the D38E mutation leads to a 240-fold decrease in  $k_{cat}$ <sup>58</sup> and a crystal structure of this mutant showed that the mutation perturbs the position of the general base by  $\sim 1 \text{ \AA}$  without introducing additional structural rearrangements.<sup>23</sup> These results suggested that the rate reduction arises predominantly from local mispositioning of the general base upon lengthening its side chain by one methylene group and allowed us to use the D38E mutation in double-mutant systems to probe the effects of other mutations on general base positioning.<sup>23</sup> In the simplest model, if a residue is important for positioning the general base, mutation of that residue will lead to a less deleterious effect in a KSI mutant with an already mispositioned general base (i.e., in the D38E mutant background) than in WT KSI with a well-positioned general base (i.e., in the Asp38 background). We previously applied this double-mutant cycle approach to assess the roles of Phe54 and Phe116, which directly interact with the Asp38 carboxylate atoms and apparently contribute to the positioning of the Asp38 carboxylate group via the anion–aromatic interactions noted above.<sup>23</sup>

To further investigate the effects of the D38E mutation, we compared molecular dynamics (MD) simulations of WT and D38E KSI (Table 1 and Figure 3). Because we have more precise structural information for bound equilenin, an intermediate analogue, than we do for the reaction substrate and product, we simulated the dynamics of the KSI-intermediate complex (see Methods). Mixed quantum mechanical/molecular mechanical

**Table 1. Average Proton Donor–Acceptor Distances (angstroms) from MD Simulations<sup>a</sup>**

enzyme	step 1	step 2
WT	2.97 (0.01)	3.66 (0.04)
A114G	2.93 (0.01)	3.82 (0.12)
P39G	4.79 (1.22)	3.92 (0.86)
P39G/V40G/S42G	5.67 (1.32)	4.54 (1.15)
S42G	2.97 (0.04)	3.70 (0.15)
V40G/S42G	3.02 (0.07)	3.53 (0.25)
D38E	4.22 (0.34)	5.29 (0.48)
D38E/A114G	3.67 (0.35)	4.03 (0.67)
D38E/P39G	5.02 (1.07)	4.14 (1.01)
D38E/P39G/V40G/S42G	6.57 (0.72)	5.35 (0.45)

<sup>a</sup>Distances are calculated between the protonated oxygen atom of the general base carboxylic acid and C4 of the steroid for step 1 and C6 of the steroid for step 2 as shown in Figure 1B. The values were averaged over two 20 ns MD trajectories and two active sites. Simulations were performed for the intermediate state of the isomerization reaction of 5-AND. Standard deviations among the four data sets are given in parentheses.



**Figure 3.** Atomic isodensity surfaces of the protonated catalytic base carboxylate group from the MD simulations. Oxygen density is colored red and hydrogen density white. In each case, the greatest 90% of the oxygen or hydrogen density is shown as a transparent isosurface, and the greatest 50% of the oxygen or hydrogen density is shown as an opaque isosurface. The MD average structures of the heavy atoms of the steroid intermediate (5-AND), the catalytic base Asp38(Glu), and the active site residues Tyr14, Tyr55, and Asp99 are shown.

(QM/MM) calculations of the catalyzed reaction could provide further insight into the impact of these mutations, but the large number of mutations in this study and the long simulation times required to achieve convergence precluded the use of these more expensive methods.

Consistent with the X-ray structures, the carboxylic acid moiety of the Glu at position 38 is displaced, with an average donor–acceptor distance  $> 1 \text{ \AA}$  larger than that with Asp at this position (Table 1). Furthermore, the general base in the D38E mutant occupies several conformations with different donor–acceptor distances that are visible in the probability distributions of these distances (Figures S1 and S2 of the Supporting Information). Root-mean-square fluctuations (rmsfs) of the side chain carbon atoms (Table 2) and atomic isodensity surfaces of the general base carboxylate (Figure 3) also illustrate more mobility of the general base than in WT. The MD results are consistent with mispositioning and a greater degree of motion of the mutant Glu carboxylate group than of the WT Asp group and

**Table 2.** Root-Mean-Square Fluctuations (angstroms) for Carbon Atoms in the Catalytic Base Side Chain<sup>a</sup>

enzyme	C <sub>α</sub>	C <sub>β</sub>	C <sub>γ</sub>
WT	0.48 (0.02)	0.52 (0.03)	0.53 (0.03)
A114G	0.51 (0.04)	0.56 (0.04)	0.58 (0.04)
P39G	0.66 (0.10)	0.81 (0.17)	1.01 (0.28)
P39G/V40G/S42G	0.53 (0.14)	0.66 (0.18)	0.92 (0.38)
S42G	0.48 (0.02)	0.52 (0.03)	0.54 (0.04)
V40G/S42G	0.51 (0.03)	0.57 (0.04)	0.62 (0.07)
enzyme	C <sub>α</sub>	C <sub>β</sub>	C <sub>γ</sub>
D38E	0.69 (0.21)	0.87 (0.30)	1.03 (0.54)
D38E/A114G	0.53 (0.10)	0.64 (0.17)	0.98 (0.40)
D38E/P39G	0.50 (0.03)	0.63 (0.11)	0.72 (0.17)
D38E/P39G/ V40G/S42G	0.69 (0.06)	0.89 (0.08)	1.21 (0.07)
			C <sub>δ</sub>

<sup>a</sup>Results obtained by averaging over two 20 ns MD trajectories and two active sites. Standard deviations among the four data sets are shown in parentheses. Note C<sub>δ</sub> is only present in species containing the D38E mutation.

support the use of this comparison in double-mutant cycles designed to probe the effects of packing and loop residues on the positioning of Asp38.

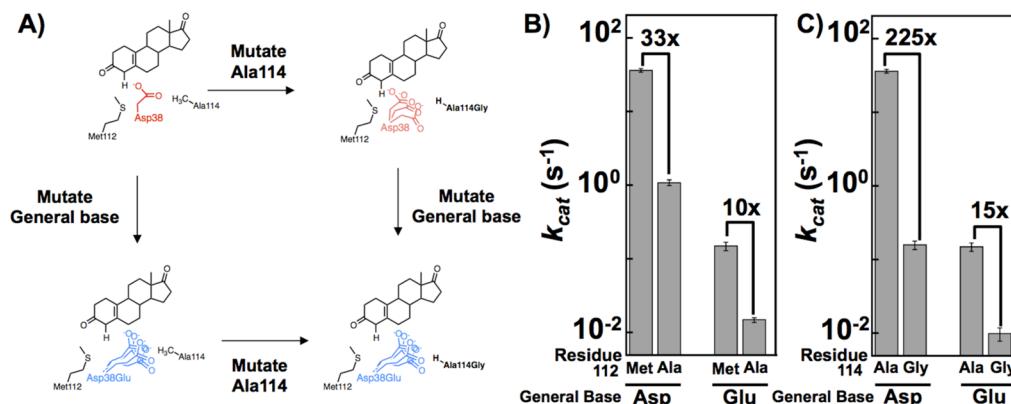
**Hydrophobic Packing with the General Base Side Chain.** Crystal structures of KSI show that the hydrophobic side chains of Met112 and Ala114 are situated within 4 Å of the general base side chain (Figure 2B), suggesting that hydrophobic packing with these side chains may restrict the conformational freedom of the general base and position it within the active site. To functionally probe this hypothesis, we used double-mutant cycles, illustrated schematically in Figure 4A for Ala114, with the M112A and A114G mutations to reduce the size of each hydrophobic side chain. The M112A and A114G mutations gave 33- and 225-fold decreases in  $k_{cat}$ , respectively (Table 3, Figure 4B,C). Throughout this study, we observe similar reductions in  $k_{cat}$  and  $k_{cat}/K_M$  (Figure S6 of the Supporting Information) and will discuss only rate reductions in  $k_{cat}$  for the remainder of the paper. We have utilized a substrate known to be limited by a chemical rather than physical step;<sup>4,59</sup> nevertheless, there are two chemical steps in the KSI reaction (Figure 1), and it is possible that the rate-limiting step changes with different mutations,<sup>60</sup> an

interesting and potentially informative possibility that is not addressed in this study.

The A114G mutational effect was greatly reduced in the D38E mutant background, from 225- to 15-fold, whereas the M112A effect was within 3-fold in both general base backgrounds. The larger effect of the Ala114 mutation with Asp than with Glu as the general base suggests that it helps to position the general base, whereas the similar effect from mutation of Met112 in both backgrounds suggests that other interactions are sufficient for near-optimal positioning of the general base. Met112 might instead contribute to catalysis via interactions with the oxyanion hole, a model that can be tested via future double-mutant cycles with oxyanion hole mutants, or to the overall stability of the active site.<sup>a</sup>

In MD simulations of the A114G mutant, we observed a nearly negligible decrease in the distance between the catalytic base hydroxyl oxygen and C4 of the steroid and a small yet potentially significant increase in the distance between the catalytic base hydroxyl oxygen and C6 of the steroid (Table 1). In addition to this change, the active site of the A114G mutant exhibits a small conformational rearrangement, relative to the WT, with the catalytic base shifted slightly in the direction of the absent methyl group in Ala114Gly (Figure 5). The remainder of the enzyme's structure is unaffected by the A114G mutation (Figure S3 of the Supporting Information), and the oxyanion hole hydrogen bonds are maintained (Table S3 of the Supporting Information). Thus, the MD results support a decrease in the level of general base function and overall catalysis because of a reorientation of the carboxylate toward the cavity created by the A114G mutation (Figure 6B) and provide no indication of an effect due to broader conformational sampling of the carboxylate group (Figure 6C). We nevertheless caution that the proton transfer reaction is likely to occur in conformers with shorter distances between the proton donor and acceptor<sup>20</sup> that are minimally sampled by equilibrium MD simulations (Figures S1 and S2 of the Supporting Information).

**Role of the Catalytic Base Loop in General Base Positioning.** As described above, the general base Asp38 in KSI is situated at the beginning of a loop that is composed of Pro39, Val40, Gly41, Ser42, Glu43, and Pro44 (Figure 2). This loop connects two strands of an antiparallel  $\beta$ -sheet and meets the geometrical definition of a type II  $\beta$ -turn.<sup>57</sup> A role for Pro39 in

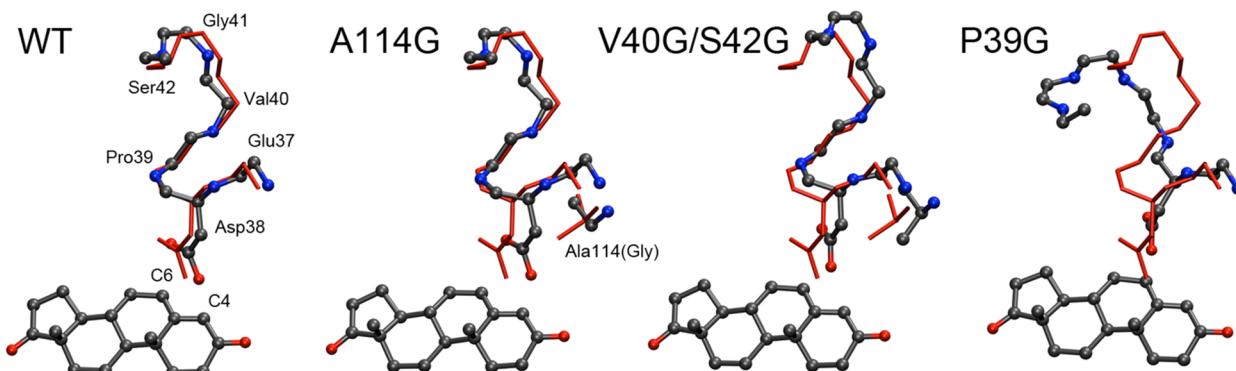


**Figure 4.** Double-mutant cycle designed to test the roles of Met112 and Ala114 in positioning the general base. (A) Mutation of Asp38 to the “mispositioned” Glu general base allows it to populate alternate conformations (vertical arrows). Mutation of Ala114 affects the positioning of the general base starting from an enzyme with a positioned general base (top horizontal arrow) but has a smaller effect for an enzyme with a mispositioned general base (bottom horizontal arrow). For the sake of clarity, only the double-mutant cycle for Ala114 is shown. Effects on  $k_{cat}$  for mutation of (B) Met112 to Ala and (C) Ala114 to Gly in KSI with a positioned Asp general base and a mispositioned D38E mutant general base. Values are from Table 3.

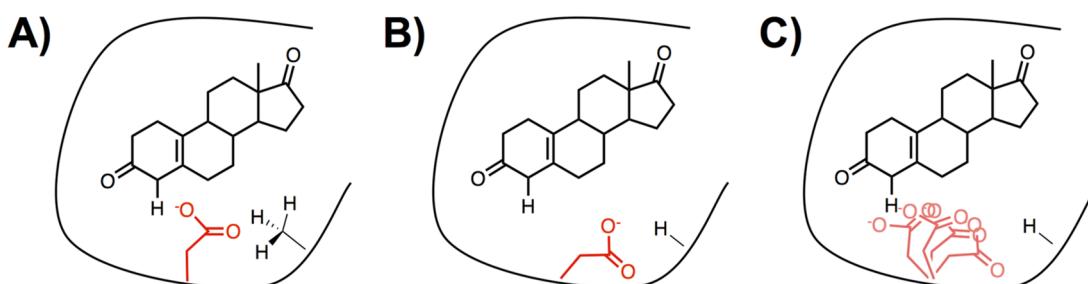
Table 3. Effects of Mutation of Met112 and Ala114 on KSI Kinetic Parameters with an Asp (WT) and Glu (D38E) General Base<sup>a</sup>

enzyme	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)	$k_{cat}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{cat}$ ratio (WT/mutant) <sup>b</sup>	$K_M$ ratio (WT/mutant) <sup>b</sup>	$k_{cat}/K_M$ ratio (WT/mutant) <sup>b</sup>
WT	36 ± 2	50 ± 4	(7.2 ± 0.3) × 10 <sup>5</sup>	(1)	(1)	(1)
M112A	1.1 ± 0.1	28 ± 6	(3.9 ± 0.5) × 10 <sup>4</sup>	33	1.8	18
A114G	(1.6 ± 0.2) × 10 <sup>-1</sup>	35 ± 4	(4.6 ± 0.3) × 10 <sup>3</sup>	225	1.4	160
enzyme	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)	$k_{cat}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{cat}$ ratio (D38E/mutant) <sup>c</sup>	$K_M$ ratio (D38E/mutant) <sup>c</sup>	$k_{cat}/K_M$ ratio (D38E/mutant) <sup>c</sup>
D38E	(1.5 ± 0.2) × 10 <sup>-1</sup>	35 ± 3	(4.3 ± 0.1) × 10 <sup>3</sup>	(1)	(1)	(1)
D38E/M112A	(1.5 ± 0.1) × 10 <sup>-2</sup>	20 ± 4	(7.5 ± 0.3) × 10 <sup>2</sup>	10	1.8	5.7
D38E/A114G	(1.0 ± 0.2) × 10 <sup>-2</sup>	19 ± 2	(5.3 ± 0.2) × 10 <sup>2</sup>	15	1.8	8.1

<sup>a</sup>Kinetics experiments were performed for the S(10)-EST substrate for which the chemical steps of the KSI reaction are rate-limiting.<sup>4,59</sup> <sup>b</sup>By definition, the ratio for WT KSI is 1, as represented by the values in parentheses. <sup>c</sup>By definition, the ratio for D38E KSI is 1, as represented by the values in parentheses.



**Figure 5.** Average structures of WT KSI and the A114G, V40G/S42G, and P39G mutants obtained from MD trajectories. The average structures of the backbone amide nitrogen,  $\alpha$ -carbon, and carbonyl carbon of residues 37–42 and the catalytic base side chain (Asp38) are depicted as color-coded balls and sticks with the corresponding atoms in the D38N-equilenin crystal structure (PDB entry 1QJG) colored red for reference. The average structure of the steroid intermediate from the MD trajectories is shown as balls and sticks, as well. In addition, the average structure of Ala114(Gly) is shown for the A114G and V40G/S42G mutants. A representative data set is shown in each case, with the remaining data sets provided in the Supporting Information.



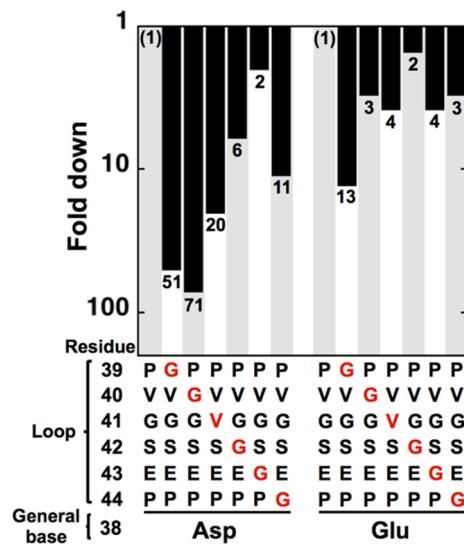
**Figure 6.** Models illustrating possible effects of mutations on the positioning of the catalytic base carboxylate group (red) in the KSI active site. (A) In WT KSI, packing interactions with Ala114 position the Asp carboxylate relative to the substrate. The cavity created by the Ala114Gly mutation allows the carboxylate to populate a single mispositioned conformation (B) or occupy multiple conformations within the active site with a subset of conformations that are mispositioned (C).

positioning the general base was suggested previously by Choi et al. on the basis of the deleterious rate effects from mutations of Pro39 to alanine or glycine, as well as a crystal structure of the P39A mutant that showed the general base carboxylate displaced by  $\sim$ 3 Å relative to the WT.<sup>61</sup> We mutated each loop residue in both the Asp38 (WT) and D38E mutant backgrounds to functionally test each for a role in general base positioning, and we made several combined mutations to probe for cooperative and collective effects.

Mutations of the loop residues Pro39 and Val40 to glycine led to 51- and 71-fold decreases in  $k_{cat}$ , respectively. Smaller rate reductions of 2–20-fold were observed for mutations of the remaining residues in the loop, G41V, S42G, E43G, and P44G (Figure 7 and Figure S7 of the Supporting Information). In the

D38E background, mutation of residues 42 and 44 led to 2–3-fold rate reductions, slightly smaller than in the WT background, suggesting only a small role of these residues in general base positioning. In contrast, mutation of Glu43 to glycine in the mispositioned D38E background gave effects that were  $\sim$ 2-fold larger than in the native Asp38 background, suggesting that Glu43 plays a slightly larger role in catalysis with the mutant Glu than with the native Asp general base.

Turning to the residues giving the largest effects, Pro39, Val40, and Gly41, we found that the rate reduction due to mutation of each of these residues was considerably smaller in the D38E background than with the WT Asp general base (Figure 7 and Figure S7 of the Supporting Information), suggesting that each of these residues plays a role in positioning the general base for



**Figure 7.** Double-mutant cycles designed to test the role of the loop residues Pro39–Pro44 in positioning the general base. Effects on  $k_{cat}$  for the mutation of loop residues in KSI with a positioned Asp38 general base and a mispositioned D38E mutant general base are shown. Values are plotted relative to WT KSI on the left-hand side and relative to the D38E mutant on the right-hand side. Data are from Table 4.

catalysis. Mutation of Pro39 to glycine in the D38E background retained a 13-fold rate decrease, apparently consistent with Pro39 also providing some positioning for the mutant Glu general base.

These kinetics results for each loop residue mirror the conservation of each position. A Basic Local Alignment Search Tool (BLAST)<sup>62,63</sup> sequence search for KSI gave 419 sequences with an Asp residue at the position equivalent to Asp38. All of these sequences had proline at position 39 and glycine at position 41; position 40 was highly conserved as a hydrophobic residue (313 Val, 51 Ile, 47 Leu, and 7 Ala residues and 1 Phe residue). Conversely, the amino acids present at positions 42–44, which had the smallest functional impacts, were variable. These observations are broadly consistent with homology studies of

$\beta$ -turns in proteins, which have shown that the analogues of positions 39 and 40 exhibit slight preferences for proline, while the analogues of positions 41 and 42 exhibit slight preferences for glycine.<sup>57,64</sup>

Overall, these results lead to the general conclusion that the loop aids in positioning the general base; i.e., the anion–aromatic interactions with Phe54 and Phe116 and the side chain packing with Ala114 are not alone sufficient to favor an optimally positioned general base carboxylate group. These results further suggest that the loop does not act as a single cooperative unit: there are different effects at different positions, and there are different levels of importance of individual residues, relative to the alternative mutated residues, in positioning the general base Asp38. Below we further analyze the effects of each mutation and the role of each loop residue.

**Pro39.** The effect of Pro39 mutations in both the WT and D38E backgrounds could result from restriction of the conformational freedom of the backbone<sup>65,66</sup> and/or from accessing conformations favored by a *cis* peptide bond. Proline is unique among the amino acids in its ability to adopt a stable *cis* isomer of its peptide bond, and the *cis* conformation is conserved in the two homologous KSIs for which structures are available.<sup>7,67</sup> In contrast, the *trans* isomer is observed in crystal structures of Pro39 mutants, suggesting that the *cis* conformation may be important for KSI activity.<sup>61</sup> To learn more about the role of Pro39, we turned to MD simulations.

Comparison of the MD simulations of WT KSI and the P39G mutant reveals a greatly perturbed average structure of the loop in the P39G mutant (Figure 5), suggesting that the remainder of the loop must compensate for the *trans* conformation at the mutant position to bring the loop back to the core  $\beta$ -strands that it connects. Furthermore, the larger rmsfs of the  $C_\alpha$  atoms of each loop residue in the P39G mutant, relative to WT (Table 5), and the broader distributions of dihedral angles for the other loop residues for the P39G mutant, relative to WT (Figure 8), are indicative of increased loop flexibility.

These results suggest that the P39G mutation increases the conformational flexibility of the backbone in the loop that bears

**Table 4. Effects of Individual Mutations of Catalytic Base Loop Residues on KSI Kinetic Parameters with an Asp (WT) and Glu (D38E) General Base<sup>a</sup>**

enzyme	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M^{-1} s^{-1}$ )	$k_{cat}$ ratio (WT/mutant) <sup>b</sup>	$K_M$ ratio (WT/mutant) <sup>b</sup>	$k_{cat}/K_M$ ratio (WT/mutant) <sup>b</sup>
WT	$36 \pm 2$	$50 \pm 4$	$(7.2 \pm 0.3) \times 10^5$	(1)	(1)	(1)
P39G	$(7.1 \pm 0.2) \times 10^{-1}$	$100 \pm 2$	$(7.1 \pm 0.2) \times 10^3$	51	0.5	100
V40G	$(5.1 \pm 0.2) \times 10^{-1}$	$40 \pm 2$	$(1.3 \pm 0.3) \times 10^4$	71	1.3	57
G41A	$2.6 \pm 0.2$	$50 \pm 3$	$(5.2 \pm 0.2) \times 10^4$	14	1.0	14
G41V	$1.8 \pm 0.4$	$40 \pm 3$	$(4.5 \pm 0.4) \times 10^4$	20	1.3	16
S42G	$6.3 \pm 0.2$	$34 \pm 3$	$(1.9 \pm 0.4) \times 10^5$	5.7	1.5	3.9
E43G	$16 \pm 0.2$	$25 \pm 4$	$(6.4 \pm 0.4) \times 10^5$	2.3	2.0	1.1
P44G	$3.4 \pm 0.3$	$50 \pm 5$	$(6.8 \pm 0.2) \times 10^4$	11	1.0	11
enzyme	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M^{-1} s^{-1}$ )	$k_{cat}$ ratio (D38E/mutant) <sup>c</sup>	$K_M$ ratio (D38E/mutant) <sup>c</sup>	$k_{cat}/K_M$ ratio (D38E/mutant) <sup>c</sup>
D38E	$(1.5 \pm 0.2) \times 10^{-1}$	$35 \pm 13$	$(4.3 \pm 0.1) \times 10^3$	(1)	(1)	(1)
D38E/P39G	$(1.2 \pm 0.5) \times 10^{-2}$	$30 \pm 9$	$(4.0 \pm 0.1) \times 10^2$	13	1.2	11
D38E/V40G	$(4.9 \pm 0.3) \times 10^{-2}$	$25 \pm 9$	$(2.0 \pm 0.2) \times 10^3$	3.1	1.4	2.2
D38E/G41V	$(4.0 \pm 0.2) \times 10^{-2}$	$35 \pm 5$	$(1.1 \pm 0.4) \times 10^3$	3.8	1.0	3.8
D38E/S42G	$(1.1 \pm 0.2) \times 10^{-1}$	$28 \pm 9$	$(3.9 \pm 0.1) \times 10^3$	1.4	1.3	1.1
D38E/E43G	$(4.0 \pm 0.2) \times 10^{-2}$	$50 \pm 10$	$(8.0 \pm 0.1) \times 10^2$	3.8	0.7	5.4
D38E/P44G	$(5.0 \pm 0.1) \times 10^{-2}$	$62 \pm 4$	$(8.1 \pm 0.1) \times 10^2$	3.0	0.6	5.3

<sup>a</sup>Kinetics experiments were performed for the 5(10)-EST substrate for which the chemical steps of the KSI reaction are rate-limiting.<sup>4,59</sup> <sup>b</sup>By definition, the ratio for WT KSI is 1, as represented by the values in parentheses. <sup>c</sup>By definition, the ratio for D38E KSI is 1, as represented by the values in parentheses.

**Table 5. Root-Mean-Square Fluctuations (angstroms) of  $C_\alpha$  Atoms for the Catalytic Base Loop Residues 38–42 Obtained from MD Simulations<sup>a</sup>**

enzyme	Asp38(Glu)	Pro39(Gly)	Val40(Gly)	Gly41	Ser42(Gly)
WT	0.48 (0.02)	0.58 (0.04)	0.54 (0.05)	1.00 (0.08)	0.72 (0.02)
A114G	0.50 (0.04)	0.60 (0.04)	0.55 (0.03)	1.01 (0.03)	0.76 (0.02)
P39G	0.66 (0.09)	1.00 (0.24)	1.60 (0.13)	2.85 (0.35)	2.04 (0.24)
P39G/V40G/S42G	0.53 (0.12)	0.85 (0.30)	1.20 (0.52)	1.72 (0.70)	1.59 (0.59)
S42G	0.48 (0.02)	0.55 (0.02)	0.53 (0.03)	0.90 (0.04)	1.26 (0.43)
V40G/S42G	0.51 (0.03)	0.75 (0.10)	0.97 (0.23)	1.28 (0.20)	1.51 (0.28)
D38E	0.69 (0.18)	0.75 (0.25)	0.62 (0.11)	0.98 (0.11)	0.90 (0.10)
D38E/A114G	0.53 (0.08)	0.67 (0.08)	0.66 (0.07)	1.18 (0.09)	0.85 (0.10)
D38E/P39G	0.49 (0.03)	0.79 (0.22)	0.90 (0.14)	1.61 (0.21)	1.28 (0.10)
D38E/P39G/V40G/S42G	0.68 (0.05)	1.04 (0.23)	1.88 (0.31)	2.33 (0.46)	1.94 (0.18)

<sup>a</sup>Results obtained by averaging over two 20 ns MD trajectories and two active sites. Standard deviations among the four data sets are shown in parentheses.

the general base, but a catalytic effect would arise only if the base itself were mispositioned or exhibited increased flexibility. The MD simulations show increased average proton donor–acceptor distances for both proton transfer positions relative to those of the WT (Table 1), and the probability distributions of these distances are broader, as well (Figures S1 and S2 of the Supporting Information), consistent with the rate reductions observed experimentally. These probability distributions and the three-dimensional atomic isodensity surfaces of the catalytic base carboxylate (Figure 3) reveal two types of conformations: a “catalytically competent” type (loosely defined), in which the carboxylic acid group is situated near C6 of the steroid, and a “catalytically incompetent” type, in which the carboxylic acid is disordered and interacts with the solvent.<sup>b</sup> The most probable distances in each type of conformation are provided in Tables S1 and S2 of the Supporting Information.

The functional and MD results together suggest that Pro39 works with the rest of the loop to position the catalytic base such that, when Pro39 is mutated to glycine, the mutant residue adopts the *trans* configuration, the remaining loop residues become more flexible, and the positioning of the general base side chain is disrupted. The *cis* conformation is ~2.5 kcal/mol less stable than the *trans* conformation for non-prolyl amino acids in solution,<sup>68</sup> providing an estimate for the upper limit of the conformational preference of the remaining loop residues of the Pro39 mutants for the *cis* conformation relative to the *trans* conformation.

**Val40.** A recent study using continuum electrostatics calculations and a structure-based algorithm to predict catalytically important residues in enzymes suggested that remote residues make only small contributions to catalysis in the homologous KSI from *Pseudomonas putida*.<sup>69</sup> However, we observe a 71-fold rate decrease with the distal V40G mutation (Figure 7).

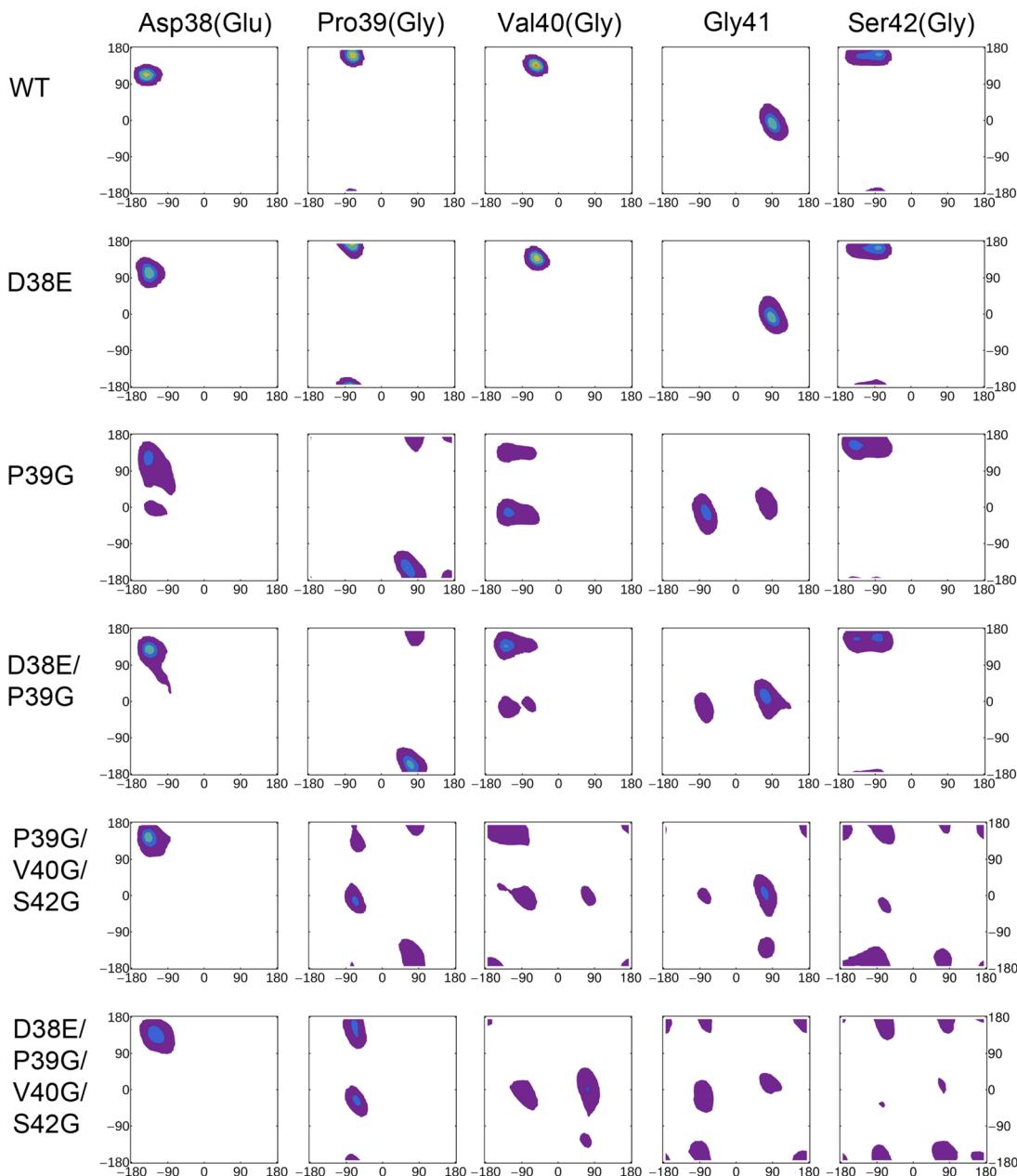
Structural inspection revealed that Val40 is situated at the interface of the KSI dimer and is positioned to pack against Val74 in a loop in the adjacent monomer (Figure 9A), raising the possibility that packing interactions extending across the dimer interface contribute to function. To test whether loop–loop packing interactions across the dimer interface could help structure the loop adjacent to Asp38 and thereby help to position the general base, we determined the effect of the V74G mutation in KSI with the structured WT loop and in a construct with multiple loop residues mutated to glycine (P39G/V40G/S42G) to ensure that loop positioning was disrupted (as will be discussed further below).

The V74G mutation decreased activity 5-fold with the loop residues intact, whereas the same mutation slightly increased activity in an enzymatic background with the loop residues mutated to glycine (Figure 9B). These data suggest that packing interactions across the dimer interface provide a modest but significant contribution to the positioning of the general base.

To gain insight into the effects of the V74G mutation and the residues with which it interacts, we examined the average structures from the MD simulations of each mutant in the residues proximal to Val74. In MD simulations of the P39G mutant (and other mutants containing the P39G mutation), we observe a conformational change in the backbone of residues 74–76 (Figure 10 and Figure S13 of the Supporting Information), whereas the average structure of residues 74–76 is well maintained in MD simulations of WT KSI and mutants not containing the P39G mutation. This observation suggests the existence of a structural link between these sites such that a perturbation in one loop affects the structure of the other across the dimer interface. On the basis of the contributions of the residues in the general base loop to positioning the general base described above, these results are consistent with packing interactions between these loops that affect each loop’s structure and thereby help to position the general base.

An alternative model for the effects of the V40G mutation is suggested by inspection of crystal structures. Val40 is situated to pack against the residues in the  $\beta$ -strand bearing Ala114, which positions the general base via hydrophobic packing interactions with its side chain as described above. Mutation of Val40 might disrupt the positioning of the general base by perturbing the packing interactions with the adjacent  $\beta$ -strand, thereby perturbing the packing interactions between Ala114 and the general base. The 71- and 5-fold rate reductions due to the V40G and V74G mutations, respectively, are smaller than the 200-fold rate reduction due to the A114G mutation, which is consistent with the rate effects of the V40G and V74G mutations acting through Ala114, whereby hydrophobic packing interactions with Val40 and secondary effects from Val74 would help position Ala114, which in turn positions the general base as described above.

To probe this alternative model, we examined the average structure from MD simulations of the V40G/S42G mutant. The structure of this mutant exhibits a conformational change in the general base side chain that is similar to the conformational change observed upon mutation of Ala114 (Figure 5). In addition to this change, the average structure of Ala114 in the V40G/S42G mutant is shifted further from the general base side

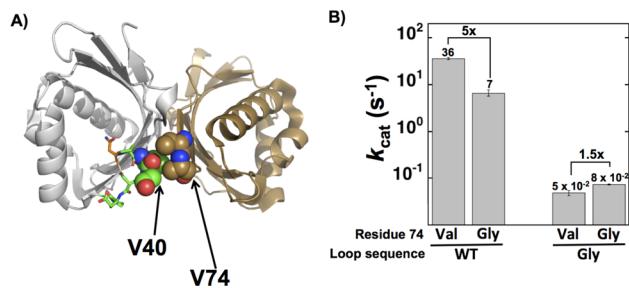


**Figure 8.** Ramachandran plots of the catalytic base loop residues Asp38(Glu), Pro39(Gly), Val40(Gly), Gly41, and Ser42(Gly) obtained from MD trajectories by calculating the probability distribution function averaged over all data sets. The  $x$ -axis corresponds to angle  $\varphi$  (degrees), and the  $y$ -axis corresponds to angle  $\psi$  (degrees). Each row corresponds to the specified mutant, and each column corresponds to the specified residue in the catalytic base loop.

chain relative to its position in the WT enzyme. These observations are consistent with the model in which mutations of the more distal Val40 and Val74 affect the positioning of the general base via perturbation of the packing interactions between the general base Asp38 and the hydrophobic side chain of Ala114. This model leads to the testable prediction that the V40G and V74G mutations will be less deleterious in the A114G background.

**Gly41.** The loop residue Gly41 was found to be conserved among KSI homologues in a BLAST sequence search and occurs at a position of a  $\beta$ -turn that is frequently glycine according to surveys of the Protein Data Bank.<sup>57,70</sup> Furthermore, KSI crystal structures indicate that the backbone dihedral angles of Gly41 are in a region of the Ramachandran plot that is unfavorable for all

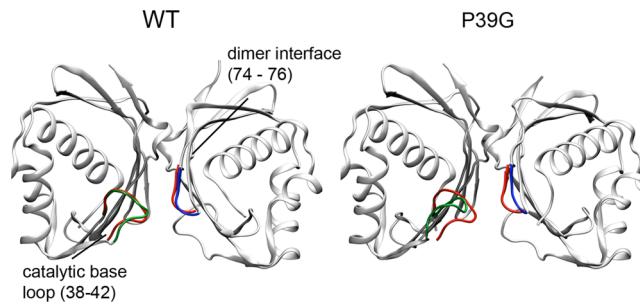
other amino acids (Figure 8). The 14- and 20-fold rate decreases observed experimentally for the G41A and G41V mutations, respectively (Table 4), support the hypothesis that adoption of a backbone conformation more favorable for glycine than for the other amino acids at position 41 is important for the loop structure, Asp38 positioning, and thus KSI activity. Furthermore, the smaller 4-fold effect of the G41V mutation in the D38E mutant background suggests that Gly41 helps to position the general base. The similar rate reductions from mutations of Gly41 to alanine and the bulkier valine provide no support for the alternative model in which unfavorable steric overlap involving the side chain is responsible for the rate reductions in these mutants.



**Figure 9.** Double-mutant cycle analysis designed to test the hypothesis that packing interactions across the dimer interface position the general base in KSI. (A) In WT KSI, Val40 of monomer A (gray) and Val74 of monomer B (gold) are situated near each other. (B) Effects on  $k_{cat}$  for mutation of Val74 to Gly with a WT loop sequence and with the loop residues mutated to glycine are shown, where the Gly loop designation refers to the P39G/V40G/S42G mutant background. Values are averages from three or more independent measurements and are from Table 6.

MD simulations of WT KSI and the D38E mutant reveal the preponderance of a single conformation of Gly41, one that is unfavorable for all other amino acids (Figure 8). We did not perform MD simulations of the Gly41 mutants, because of the unavailability of an appropriate crystal structure, but in simulations of mutants with P39G, multiple conformations of Gly41 are observed, including regions of Ramachandran space that are accessible to all other amino acids and not unique to glycine residues (Figure 8). Gly41 appears to aid general base positioning and catalysis by allowing a loop conformation that requires occupancy of a region of Ramachandran space that is inaccessible to other amino acids.

**Ser42, Glu43, and Pro44.** The observation, described above, that the overall structure of the loop is important for general base positioning and KSI activity might suggest significant contributions from all of the residues in the loop. In particular, crystal structures of KSI illustrate that the side chain hydroxyl group of Ser42 is positioned to hydrogen bond with the backbone carbonyl group of Pro39, suggesting that this interaction could stabilize the native loop structure. Nevertheless, the S42G mutation led to an only 6-fold rate decrease (Figure 7). The results from MD simulations of the S42G mutant are virtually indistinguishable from those of WT KSI in all analyses performed: donor–acceptor distances (Table 1), atomic isodensity surfaces (Figure S4 of the Supporting Information), and Ramachandran plots (Figure S5 of the Supporting Information). These similarities in the MD simulations are consistent with the small experimental rate reduction.



**Figure 10.** Ribbon diagrams of the average structure of WT KSI (left) and the P39G mutant (right), both colored gray, highlighting the conformational changes in the catalytic base loop residues (38–43, green) and the proximal residues across the dimeric interface (74–76, blue). The backbone structures of these two regions in the D38N-equilenin crystal structure (PDB entry 1QJG) are colored red for reference. Structures were aligned by minimization of the rmsd of the C<sub>α</sub> atoms in the dimer.

Mutations of the remaining residues comprising the loop, Glu43 and Pro44, led to rate reductions of 2- and 11-fold, respectively, for  $k_{cat}$  (Figure 7). Surprisingly, mutation of Glu43 to glycine in the mispositioned D38E background gave effects that were ~2-fold larger than in the native Asp38 background. Alternate conformations may be allowed with E43G that increase the probability of reaction utilizing the alternative Glu38 general base. Future tests of our understanding might involve attempts to reoptimize catalysis with Glu at position 38 or at other positions in the loop.<sup>71–73</sup>

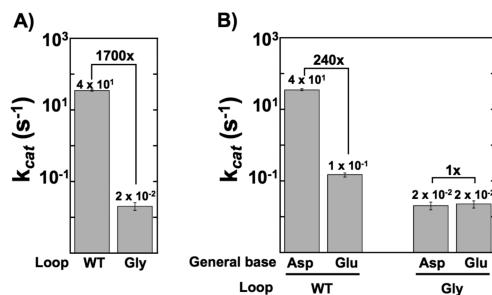
**Multiple Glycine Mutations for Evaluating the Contribution of the Full Loop to General Base Positioning.** To obtain an upper bound on the catalytic effect of the loop, we mutated all of the loop residues to a polyglycine loop and conducted a double-mutant cycle analysis as described above. Mutation of the loop to an all-glycine sequence with Asp38 present led to a 1700-fold rate decrease (Figure 11A), and mutation to Glu38 had no additional effect (Figure 11B), indicating that any preferential positioning of Asp at position 38 had been eliminated.

The 1700-fold effect from conversion to the all-glycine loop (Figure 11A) is larger than the 240-fold effect from mutating Asp to Glu with the WT loop (Figure 11B) and constitutes our best estimate for the full contribution of the loop to positioning, relative to an unpositioned loop. The smaller effect from the D38E mutation alone, with the WT loop present, would be expected if the WT loop helped position Glu to function as the general base, relative to an all-glycine loop. Alternatively, the larger effect on activity from conversion to the all-glycine loop with Asp38 present could arise from secondary structural effects from mutation of the loop that impair catalysis in other ways.

**Table 6. Effects of Mutation of Val74, a Residue Situated across the Dimer Interface from the Structured Loop, in WT KSI and in a KSI Mutant with Multiple Residues in the Loop Mutated to Glycine<sup>a</sup>**

enzyme	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)	$k_{cat}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{cat}$ ratio (WT/mutant) <sup>b</sup>	$K_M$ ratio (WT/mutant) <sup>b</sup>	$k_{cat}/K_M$ ratio (WT/mutant) <sup>b</sup>
WT	$36 \pm 2$	$50 \pm 4$	$(7.2 \pm 0.3) \times 10^5$	(1)	(1)	(1)
V74G	$6.7 \pm 2$	$35 \pm 4$	$(1.9 \pm 0.3) \times 10^5$	5.4	1.4	3.8
P39G/V40G/S42G	$(4.8 \pm 0.6) \times 10^{-2}$	$55 \pm 11$	$(8.7 \pm 0.2) \times 10^2$	750	0.9	830
P39G/V40G/S42G/ V74G	$(7.5 \pm 0.2) \times 10^{-2}$	$35 \pm 2$	$(2.1 \pm 0.2) \times 10^3$	480	1.4	340

<sup>a</sup>Kinetics experiments were performed for the 5(10)-EST substrate for which the chemical steps of the KSI reaction are rate-limiting.<sup>4,59</sup> <sup>b</sup>By definition, the ratio for WT KSI is 1, as represented by the values in parentheses.



**Figure 11.** Double-mutant cycle analysis designed to test the collective role of the catalytic base loop in positioning the general base in the KSI active site. (A) Effects on  $k_{\text{cat}}$  for mutation of the residues in the loop to glycine, where the Gly loop designation refers to the P39G/V40G/S42G/E43G/P44G mutant. (B) Effects on  $k_{\text{cat}}$  for mutation of the Asp38 general base to Glu with a WT loop sequence and with the residues in the loop mutated to glycine. Values are from Table 7.

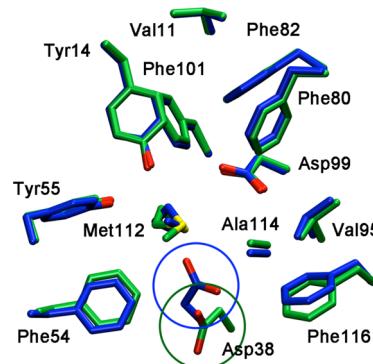
However, the similar  $K_M$  values for WT KSI and the all-glycine loop mutant (Table 7) provide no evidence for binding effects due to the mutations, and several additional observations, described below, also provide no indication of structural rearrangements beyond the general base position. We therefore suggest that the 1700-fold effect provides a reasonable measure of the overall positioning effect from the loop, relative to the unstructured all-glycine loop and in the context of the rest of KSI being folded.

To directly test whether mutating multiple loop residues to glycine introduces further structural rearrangements, we determined the crystal structures of the P39G/V40G/S42G (PDB entry 3OV4) and D38E/P39G/V40G/S42G (PDB entry 3NM2) mutants at 1.8 and 1.9 Å resolution, respectively. Refinement details are given in Table S4 of the Supporting Information. The catalytic rates of these mutants are within 2-fold of the catalytic rates of the all-glycine loop mutants, and the D38E mutation has a negligible rate effect in both the P39G/V40G/S42G and all-glycine mutant backgrounds (Table 7), suggesting that these structures are representative of the all-glycine loop mutants.

The overall structures of these mutants can be superimposed with that of WT KSI [PDB entry 8CHO<sup>74</sup>] with  $C_\alpha$  rmsds of 0.71

and 0.52 Å, respectively, for a single monomer, excluding residues 38–42, suggesting that these mutations do not cause significant rearrangements outside of the loop. We observed poor electron density for the loop backbone atoms in both structures, suggesting that the loop is disordered, as would be expected.

This disorder observed in the loop residues appears to affect general base positioning. In the P39G/V40G/S42G mutant, the loop residues have poorly defined electron density, the refined position of the Asp38 carboxylate group in all four chains of the asymmetric unit is displaced by ~2.5 Å relative to that of the WT (Figure 12), and the normalized  $B$  factors in the Asp38 side chain are larger than those in the WT enzyme (Table S5 of the Supporting Information).<sup>75</sup> Furthermore, a distinct refined position of this carboxylate group is found in each asymmetric



**Figure 12.** Superposition of the 1.8 Å structure of the P39G/V40G/S42G mutant determined herein (PDB entry 3OV4, carbon atoms colored green) and the previously determined 2.3 Å structure of WT KSI (PDB entry 8CHO, carbon atoms colored blue). In both structures, oxygen atoms are colored red and sulfur atoms yellow. The crystal structure of the P39G/V40G/S42G mutant shows that the base is mispositioned relative to WT KSI in all four chains in the asymmetric unit. The overall root-mean-square deviation between the two structures for backbone atoms is 0.6 Å. For the sake of clarity, only the A chain is shown for the P39G/V40G/S42G structure. The Asp38 carboxylate group is displaced by ~2.5 Å in the mutant compared to WT KSI, as indicated by the green and blue circles, respectively.

**Table 7. Effects of Mutating Multiple Loop Residues to Glycine on KSI Kinetic Parameters with an Asp (WT) and Glu (D38E) General Base<sup>a</sup>**

enzyme	$k_{\text{cat}}$ (s <sup>-1</sup> )	$K_M$ (μM)	$k_{\text{cat}}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{\text{cat}}$ ratio (WT/mutant) <sup>b</sup>	$K_M$ ratio (WT/mutant) <sup>b</sup>	$k_{\text{cat}}/K_M$ ratio (WT/mutant) <sup>b</sup>
WT	36 ± 2	50 ± 4	(7.2 ± 0.3) × 10 <sup>5</sup>	(1)	(1)	(1)
P39G	(7.1 ± 0.6) × 10 <sup>-1</sup>	100 ± 11	(7.1 ± 0.2) × 10 <sup>3</sup>	51	0.5	100
P39G/V40G	(8.6 ± 0.3) × 10 <sup>-2</sup>	55 ± 12	(1.6 ± 0.3) × 10 <sup>3</sup>	420	0.9	460
P39G/V40G/S42G	(4.8 ± 0.6) × 10 <sup>-2</sup>	55 ± 11	(8.7 ± 0.2) × 10 <sup>2</sup>	750	0.9	830
P39G/V40G/S42G/E43G	(3.9 ± 0.1) × 10 <sup>-2</sup>	35 ± 8	(1.1 ± 0.4) × 10 <sup>3</sup>	920	1.4	650
P39G/V40G/S42G/E43G/P44G	(2.1 ± 0.5) × 10 <sup>-2</sup>	50 ± 16	(4.2 ± 0.4) × 10 <sup>2</sup>	1700	1.0	1700
enzyme	$k_{\text{cat}}$ (s <sup>-1</sup> )	$K_M$ (μM)	$k_{\text{cat}}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{\text{cat}}$ ratio (D38E/mutant) <sup>c</sup>	$K_M$ ratio (D38E/mutant) <sup>c</sup>	$k_{\text{cat}}/K_M$ ratio (D38E/mutant) <sup>c</sup>
D38E	(1.5 ± 0.2) × 10 <sup>-1</sup>	35 ± 13	(4.3 ± 0.1) × 10 <sup>3</sup>	(1)	(1)	(1)
D38E/P39G/V40G/S42G	(4.0 ± 0.2) × 10 <sup>-2</sup>	34 ± 9	(1.2 ± 0.2) × 10 <sup>3</sup>	3.8	1.0	3.6
D38E/P39G/V40G/S42G/E43G/P44G	(2.3 ± 0.5) × 10 <sup>-2</sup>	68 ± 7	(3.4 ± 0.3) × 10 <sup>2</sup>	6.5	0.51	13

<sup>a</sup>Kinetics experiments were performed for the 5(10)-EST substrate for which the chemical steps of the KSI reaction are rate-limiting.<sup>4,59</sup> <sup>b</sup>By definition, the ratio for WT KSI is 1, as represented by the values in parentheses. <sup>c</sup>By definition, the ratio for D38E KSI is 1, as represented by the values in parentheses.

unit of this structure. These structural data, similar to the functional data presented above, suggest that the all-glycine loop disrupts positioning of the Asp38 general base.

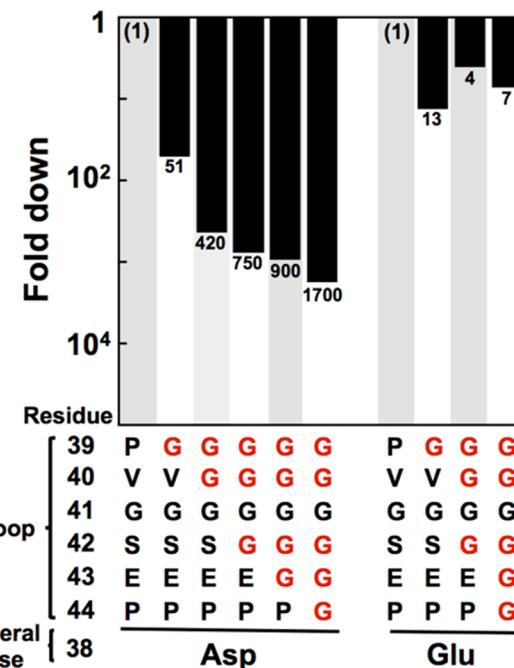
To further investigate the positioning of the general base in the polyglycine loop mutants, we performed MD simulations of the P39G/V40G/S42G and D38E/P39G/V40G/S42G mutants and compared these results to those of the single mutants examined above. The calculated rmsfs of  $C_{\alpha}$  atoms in the loop (Table 5) suggest that the rmsfs of the polyglycine mutants are similar to or smaller than those of the single glycine mutant, a result that is somewhat surprising. Ramachandran plots of the residues in the loop (Figure 8), however, illustrate that the polyglycine loop mutants P39G/V40G/S42G and D38E/P39G/V40G/S42G exhibit many local minima in residues 39–42, consistent with the presence of a disordered loop. We caution that while these distributions illustrate an increase in the number of allowed conformations for the polyglycine mutants, they may not be quantitatively accurate on the basis of the limitations of equilibrium MD simulations.

The increased flexibility observed in MD simulations of the loop due to polyglycine mutation would impact the rate only if it affects the reactive interface of the general base and the substrate. The average donor–acceptor distances from these simulations are larger than those of the single-glycine mutants and significantly larger than those observed in the WT (Table 1). The distance distributions are slightly broader than those of the single mutants, as well (Figures S1 and S2 of the Supporting Information). These results support the presence of an increased level of disorder in the general base side chain and are consistent with the experimentally observed rate reductions (Table 7). The slightly broader configuration space sampled by the general base can be visualized in the atomic isodensity surfaces (Figure 3).

Finally, we asked whether the loop residues act independently or cooperatively to position the general base. In a model of full cooperativity of all of the loop residues, any mutation in the loop would disrupt the positioning of the general base. This was ruled out by the data presented above showing that single mutations in different loop residues gave different effects (Figure 7). Figure 13 shows the results of successively mutating loop residues to Gly. The increasing effect from adding more residues again argues against the unlikely fully cooperative model. On the other hand, the effect from mutation to the all-glycine loop of 1700-fold is much smaller than that predicted from a model of fully independent contributions from each residue. Such a model predicts multiplicative effects from the individual mutations and an effect of  $10^{5.7}$ -fold (Table 4). Thus, an intermediate model is supported in which the residue identity at individual loop positions affects the contributions from certain other residues. Accurately describing the detailed nature of the partially cooperative interactions among these loop residues, as well as designing alternative yet effective loop sequences, provides future challenges for both theory and experiment.

## CONCLUSIONS AND IMPLICATIONS

We draw enzyme mechanisms showing so-called catalytic residues, but we know that these residues must be placed within the context of a folded protein. Understanding the functional interrelationships between the catalytic groups and their protein surroundings is an important and difficult current challenge in enzymology. Developments in this area have implications for understanding how enzymes have evolved and for developing effective strategies for designing protein and nonprotein catalysts with activities and specificities that rival those of naturally



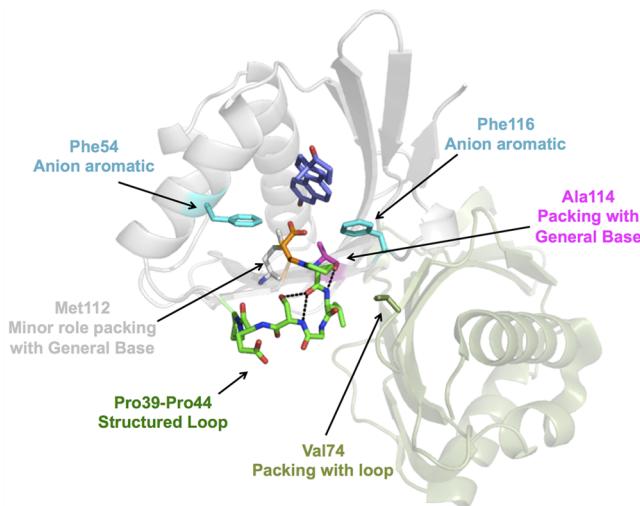
**Figure 13.** Effects of mutating multiple loop residues to glycine on  $k_{\text{cat}}$  in KSI with a positioned Asp38 general base and a mispositioned Glu (D38E mutant) general base. Values are plotted relative to that of WT KSI on the left-hand side and relative to that of the D38E mutant on the right-hand side. Data are from Table 7.

occurring enzymes. Our results reveal interrelationships involving and surrounding a highly efficient general base within an enzyme active site and provide valuable benchmarks to facilitate the development of general and quantitative models for these effects.

Prior work provided strong evidence of anion–aromatic interactions between the Asp38 carboxylate oxygen atoms and two active site Phe residues, and such interactions appear to be a common alternative to hydrogen bond networks for positioning catalytic residues.<sup>23</sup> Packing interactions of Ala114 but not Met112 with the Asp38 side chain contribute to its positioning and function. These disparate observations underscore the need to supplement structural studies with functional analysis, and double-mutant cycles as conducted herein are particularly incisive as functional probes.

The loop that begins with the general base (Figure 14) also contributes to its positioning, and we have uncovered catalytic contributions as large as  $10^3$ -fold from this loop. From another perspective, the anion–aromatic and packing interactions are insufficient to optimally position the general base, a conclusion that could not have been drawn from structural inspection alone. Further investigation of the loop contributions, via functional analysis, structural determination, and simulations, revealed differential contributions from individual loop residues and further suggested contributions both from disfavoring alternative, noncatalytic conformations and from limiting the mobility of the general base.

The residues of the loop bearing the general base in KSI exhibit partial cooperativity [i.e., mutations of different loop residues have different effects, and mutations of multiple residues have effects beyond mutation of the first (Figure 13)], but these effects fall far short of exhibiting energetic independence. Structurally, given the limited packing of protein loops relative to their interiors, we might generally expect multistate structural



**Figure 14.** Multiple interactions position the general base in the KSI active site. The structural model shows the residues that have been mutated as sticks. Residues are colored according to the interaction suggested from the functional and structural data to position the general base.

behavior of loops, with multiple conformational states explored and mutations affecting the “shape” of the conformational energy landscape, in contrast to a single dominant folded–unfolded transition as in a two-state, fully cooperative process.

Protein sectors have been suggested to evolve and function together,<sup>76,77</sup> and evolution has “mixed and matched” domains, as in dehydrogenases that use the same NAD(P)H cofactor but reduce different substrates.<sup>78,79</sup> The KSI loop adjoining Asp38 appears to position this residue to act as a general base without influencing substrate binding or the neighboring oxyanion hole. Identifying functional regions such as this may ultimately define the extent of interactions needed to create a functional catalytic motif, whether through evolution or design, and may allow us to independently alter substrate specificity and/or mutate the catalytic residue into another catalytically active group to perform distinct chemistries.

Understanding protein substrates, the conformational ensembles that describe them, and how such ensembles can be shaped and limited by engineering and by evolution to allow specificity and catalysis present major current challenges that will likely require a tight feedback loop between experiment and computation. Here we have begun this interaction focusing on the groups interacting with KSI’s general base, and new predictions arise from the MD simulations conducted herein. Nevertheless, we emphasize the need, moving forward, to further develop computational methods that can efficiently provide extensive quantitative and nontrivial predictions that can subsequently and independently be experimentally tested.

## ASSOCIATED CONTENT

### Supporting Information

Data from MD simulations, including probability distributions of donor–acceptor distances, comparison of hydrogen bonding in the oxyanion hole, overlaid ribbon diagrams of the average structures of WT KSI and the A114G mutant, data for the S42G mutant, and data from individual trajectories; refinement information from X-ray crystallography and comparison of normalized *B* factors from crystal structures; and comparison of

the effects of mutations on  $k_{cat}$  and  $k_{cat}/K_M$  values. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: shs3@illinois.edu. Phone: (217) 300-0335.  
\*E-mail: herschla@stanford.edu. Phone: (650) 723-9442.

### Funding

This work was supported by National Institutes of Health Grant GM056207 (S.H.-S.) and National Science Foundation Grant MCB1121778 (D.H.). Portions of this research were conducted at the Stanford Synchrotron Radiation Lightsource (SSRL), a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including Grant P41GM103393).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We gratefully acknowledge helpful discussions with Dr. Alexander Soudackov.

## ABBREVIATIONS

KSI, ketosteroid isomerase; MD, molecular dynamics; 5(10)-EST, 5(10)-estrene-3,17-dione; WT, wild-type; 5-AND, 5-androstene-3,17-dione; rmsd, root-mean-square deviation; rmsf, root-mean-square fluctuation; BLAST, Basic Local Alignment Search Tool; EQU, equilenin.

## ADDITIONAL NOTES

<sup>a</sup>In cases where similar mutational effects are observed in the WT and D38E backgrounds, it is also possible that the residue in question contributes similarly in positioning the D38 and D38E side chains for general base function. This alternative model is not supported by the evidence noted above that Glu positioning is different from that of Asp and does not utilize anion–aromatic interactions with Phe56 and Phe116.

<sup>b</sup>The proton transfer reaction will occur at distances significantly shorter than those sampled in equilibrium MD simulations. Here, we define catalytically competent conformations as those with proton donor–acceptor distances similar to those observed in simulations of WT KSI, whereas catalytically incompetent conformations are those with large donor–acceptor distances (>4 Å).

## REFERENCES

- (1) Pollack, R. M. (2004) Enzymatic mechanisms for catalysis of enolization: Ketosteroid isomerase. *Bioorg. Chem.* 32, 341–353.
- (2) Schwans, J. P., Kraut, D. A., and Herschlag, D. (2009) Determining the Catalytic Role of Remote Substrate Binding Interactions in Ketosteroid Isomerase. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14271–14275.
- (3) Kim, D.-H., Jang, D. S., Nam, G. H., Choi, G., Kim, J.-S., Ha, N.-C., Kim, M.-S., Oh, B.-H., and Choi, K. Y. (2000) Contribution of the hydrogen-bond network involving a tyrosine triad in the active site to the structure and function of a highly proficient ketosteroid isomerase from *Pseudomonas putida* Biotype B. *Biochemistry* 39, 4581–4589.
- (4) Hawkinson, D. C., Eames, T. C. M., and Pollack, R. M. (1991) Energetics of 3-oxo- $\Delta^5$ -steroid isomerase: Source of the catalytic power of the enzyme. *Biochemistry* 30, 10849–10858.

- (5) Choi, G., Ha, N.-C., Kim, M.-S., Hong, B.-H., Oh, B.-H., and Choi, K. Y. (2001) Pseudoreversion of the catalytic activity of Y14F by the additional substitution(s) of tyrosine with phenylalanine in the hydrogen bond network of  $\Delta^5$ -3-ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochemistry* 40, 6828–6835.
- (6) Choi, G., Ha, N.-C., Kim, S. W., Kim, D.-H., Park, S., Oh, B.-H., and Choi, K. Y. (2000) Asp-99 donates a hydrogen bond not to Tyr-14 but to the steroid directly in the catalytic mechanism of  $\Delta^5$ -3-ketosteroid isomerase from *Pseudomonas putida* Biotype B. *Biochemistry* 39, 903–909.
- (7) Wu, Z. R., Ebrahimian, S., Zawrotny, M. E., Thorneburg, L. D., Perez-Alvarado, G. C., Brothers, P., Pollack, R. M., and Summers, M. F. (1997) Solution structure of 3-oxo- $\Delta^5$ -steroid isomerase. *Science* 276, 415–418.
- (8) Kraut, D. A., Sigala, P. A., Pybus, B., Liu, C. W., Ringe, D., Petsko, G. A., and Herschlag, D. (2006) Testing electrostatic complementarity in enzymatic catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol.* 4, 501–519.
- (9) Kraut, D. A., Sigala, P. A., Fenn, T. D., and Herschlag, D. (2010) Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1960–1965.
- (10) Mildvan, A. S., Massiah, M. A., Harris, T. K., Marks, G. T., Harrison, D. H. T., Viragh, C., Reddy, P. M., and Kovach, I. M. (2002) Short, strong hydrogen bonds on enzymes: NMR and mechanistic studies. *J. Mol. Struct.* 615, 163–175.
- (11) Frey, P. A. (2001) Strong Hydrogen Bonding in Molecules and Enzymatic Complexes. *Magn. Reson. Chem.* 39, S190–S198.
- (12) Petrounia, I. P., Blotny, G., and Pollack, R. M. (2000) Binding of 2-naphthols to D38E mutants of 3-oxo- $\Delta^5$ -steroid isomerase: Variation of ligand ionization state with the nature of the electrophilic component. *Biochemistry* 39, 110–116.
- (13) Kamerlin, S. C. L., Sharma, P. K., Chu, Z. T., and Warshel, A. (2010) Ketosteroid isomerase provides further support for the idea that enzymes work by electrostatic preorganization. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4075–4080.
- (14) Warshel, A., Sharma, P. K., Chu, Z. T., and Aqvist, J. (2007) Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: Phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* 46, 1466–1476.
- (15) Park, H., and Merz, K. M., Jr. (2003) Molecular dynamics and quantum chemical studies on the catalytic mechanism of  $\Delta^5$ -3-ketosteroid isomerase: The catalytic diad versus the cooperative hydrogen bond mechanism. *J. Am. Chem. Soc.* 125, 901–911.
- (16) Mazumder, D., Kahn, K., and Bruice, T. C. (2003) Computational study of ketosteroid isomerase: Insights from molecular dynamics simulation of enzyme bound substrate and intermediate. *J. Am. Chem. Soc.* 125, 7553–7561.
- (17) Kim, K. S., Oh, K. S., and Lee, J. Y. (2000) Catalytic role of enzymes: Short strong H-bond-induced partial proton shuttles and charge redistributions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6373–6378.
- (18) Feierberg, I., and Aqvist, J. (2002) Computational modeling of enzymatic keto-enol isomerization reactions. *Theor. Chem. Acc.* 108, 71–84.
- (19) Chakravorty, D. K., and Hammes-Schiffer, S. (2010) Impact of mutation on proton transfer reactions in ketosteroid isomerase: Insights from molecular dynamics simulations. *J. Am. Chem. Soc.* 132, 7549–7555.
- (20) Chakravorty, D. K., Soudakov, A. V., and Hammes-Schiffer, S. (2009) Hybrid quantum/classical molecular dynamics simulations of the proton transfer reactions catalyzed by ketosteroid isomerase: Analysis of hydrogen bonding, conformational motions, and electrostatics. *Biochemistry* 48, 10608–10619.
- (21) Hanoian, P., Sigala, P. A., Herschlag, D., and Hammes-Schiffer, S. (2010) Hydrogen bonding in the active site of ketosteroid isomerase: Electronic inductive effects and hydrogen bond coupling. *Biochemistry* 49, 10339–10348.
- (22) Schwans, J. P., Sunden, F., Gonzalez, A., Tsai, Y., and Herschlag, D. (2011) Evaluating the Catalytic Contribution from the Oxyanion Hole in Ketosteroid Isomerase. *J. Am. Chem. Soc.* 133, 20052–20055.
- (23) Schwans, J. P., Sunden, F., Lassila, J. K., Gonzalez, A., Tsai, Y., and Herschlag, D. (2013) Use of anion-aromatic interactions to position the general base in the ketosteroid isomerase active site. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11308–11313.
- (24) Pattern, P. A., Gray, N. S., Yang, P. L., Marks, C. B., Wedemayer, G. J., Boniface, J. J., Stevens, R. C., and Schultz, P. G. (1996) The immunological evolution of catalysis. *Science* 271, 1086–1091.
- (25) Hedstrom, L. (2002) Serine Protease Mechanism and Specificity. *Chem. Rev.* 102, 4501–4523.
- (26) Kraut, D. A., Carroll, K. S., and Herschlag, D. (2003) Challenges in Enzyme Mechanism and Energetics. *Annu. Rev. Biochem.* 72, 517–571.
- (27) Hunt, J. A., Ahmed, M., and Fierke, C. A. (1999) Metal Binding Specificity in Carbonic Anhydrase Is Influenced by Conserved Hydrophobic Core Residues. *Biochemistry* 38, 9054–9062.
- (28) Ataie, N. J., Hoang, Q. Q., Zahndiser, M. P. D., Tu, Y., Milne, A., Petsko, G. A., and Ringe, D. (2008) Zinc Coordination Geometry and Ligand Binding Affinity: The Structural and Kinetic Analysis of the Second-Shell Serine 228 Residue and the Methionine 180 Residue of the Aminopeptidase from *Vibrio proteolyticus*. *Biochemistry* 47, 7673–7683.
- (29) Oue, S., Okamoto, A., Yano, T., and Kagamiyama, H. (1999) Redesigning the Substrate Specificity of an Enzyme by Cumulative Effects of the Mutations of Non-active Site Residues. *J. Biol. Chem.* 274, 2344–2349.
- (30) Lassila, J. K., Keeffe, J. R., Kast, P., and Mayo, S. L. (2007) Exhaustive Mutagenesis of Six Secondary Active-Site Residues in *Escherichia coli* Chorismate Mutase Shows the Importance of Hydrophobic Side Chains and a Helix N-Capping Position for Stability and Catalysis. *Biochemistry* 46, 6883–6891.
- (31) Benz-Moy, T. L., and Herschlag, D. (2011) Structure–Function Analysis from the Outside In: Long-Range Tertiary Contacts in RNA Exhibit Distinct Catalytic Roles. *Biochemistry* 50, 8733–8755.
- (32) Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.
- (33) Soltis, S. M., Cohen, A. E., Deacon, A., Eriksson, T., Gonzalez, A., McPhillips, S., Chui, H., Dunten, P., Hollenbeck, M., Mathews, I., Miller, M., Moorhead, P., Phizackerley, R. P., Smith, C., Song, J., van dem Bedem, H., Ellis, P., Kuhn, P., McPhillips, T., Sauter, N., Sharp, K., Tsiba, I., and Wolf, G. (2008) New paradigm for macromolecular crystallography experiments at SSRL: Automated crystal screening and remote data collection. *Acta Crystallogr. D64*, 1210–1221.
- (34) Kabsch, W. (2010) XDS. *Acta Crystallogr. D66*, 125–132.
- (35) Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
- (36) Trapani, S., and Navaza, J. (2008) AMoRe: Classical and modern. *Acta Crystallogr. D64*, 11–16.
- (37) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D67*, 355–367.
- (38) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D53*, 240–255.
- (39) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
- (40) Cho, H.-S., Ha, N.-C., Choi, G., Kim, H.-J., Lee, D., Oh, K. S., Kim, K. S., Lee, W., Choi, K. Y., and Oh, B.-H. (1999) Crystal structure of  $\Delta^5$ -3-ketosteroid isomerase from *Pseudomonas testosteroni* in complex with equilenin settles the correct hydrogen bonding scheme for transition state stabilization. *J. Biol. Chem.* 274, 32863–32868.
- (41) Kim, D.-H., Nam, G. H., Jang, D. S., Choi, G., Joo, S., Kim, J.-S., Oh, B.-H., and Choi, K. Y. (1999) Roles of Active Site Aromatic Residues in Catalysis by Ketosteroid Isomerase from *Pseudomonas putida* Biotype B. *Biochemistry* 38, 13810–13819.

- (42) Hanoian, P., and Hammes-Schiffer, S. (2011) Water in the Active Site of Ketosteroid Isomerase. *Biochemistry* 50, 6689–6700.
- (43) Xiang, J. Z., and Honig, B. (2002) JACKAL: A Protein Structure Modeling Package, Columbia University and Howard Hughes Medical Institute, New York.
- (44) Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H<sup>++</sup>: A server for estimating pK<sub>a</sub>s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33, W368–W371.
- (45) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197.
- (46) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* 65, 712–725.
- (47) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79, 926–935.
- (48) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97, 10269–10280.
- (49) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general Amber force field. *J. Comput. Chem.* 25, 1157–1174.
- (50) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092.
- (51) Essman, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593.
- (52) Hess, B. (2008) PLINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* 4, 116–122.
- (53) Nose, S. (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52, 255–268.
- (54) Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697.
- (55) Parrinello, M., and Rahman, A. (1980) Crystal Structure and Pair Potentials: A Molecular Dynamics Study. *Phys. Rev. Lett.* 45, 1196–1199.
- (56) Hess, B., Kutzner, C., Van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–437.
- (57) Wilmot, C. M., and Thornton, J. M. (1988) Analysis and prediction of the different types of  $\beta$ -turn in proteins. *J. Mol. Biol.* 203, 221–232.
- (58) Hawkinson, D. C., Pollack, R. M., and Ambulos, N. P., Jr. (1994) Evaluation of the internal equilibrium constant for 3-oxo- $\Delta^5$ -steroid isomerase using the D38E and D38N mutants: The energetic basis for catalysis. *Biochemistry* 33, 12172–12183.
- (59) Pollack, R. M., Bantia, S., Bounds, P. L., and Koffman, B. M. (1986) pH dependence of the kinetic parameters for 3-oxo- $\Delta^5$ -steroid isomerase. Substrate catalysis and inhibition by (3S)-spiro[5- $\alpha$ -androstane-3,2-oxiran]-17-one. *Biochemistry* 25, 1905–1911.
- (60) Thornburg, L. D., Goldfeder, Y. R., Wilde, T. C., and Pollack, R. M. (2001) Selective catalysis of elementary steps by Asp-99 and Tyr-14 of 3-Oxo- $\Delta^5$ -Steroid isomerase. *J. Am. Chem. Soc.* 123, 9912–9913.
- (61) Nam, G. H., Cha, S.-S., Yun, Y. S., Oh, Y. H., Hong, B. H., Lee, H.-S., and Choi, K. Y. (2003) The Conserved cis-Pro39 Residue Plays a Crucial Role in the Proper Positioning of the Catalytic Base Asp38 in Ketosteroid Isomerase from *Comamonas testosteroni*. *Biochem. J.* 375, 297–305.
- (62) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- (63) Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- (64) Kaur, H., and Raghava, G. P. S. (2003) Prediction of  $\beta$ -turns in proteins from multiple alignment using neural network. *Protein Sci.* 12, 627–634.
- (65) MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation. *J. Mol. Biol.* 218, 397–412.
- (66) Ho, B., and Brasseur, R. (2005) The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* 5, 14.
- (67) Kim, S. W., Cha, S.-S., Cho, H.-S., Kim, J.-S., Ha, N.-C., Cho, M.-J., Joo, S., Kim, K. K., Choi, K. Y., and Oh, B.-H. (1997) High-resolution crystal structures of  $\Delta^5$ -3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry* 36, 14030–14036.
- (68) Jorgensen, W. L., and Gao, J. (1988) Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution. *J. Am. Chem. Soc.* 110, 4212–4216.
- (69) Somarowthu, S., Brodin, H. R., D'Aquino, J. A., Ringe, D., Ondrechen, M. J., and Beuning, P. J. (2011) A Tale of Two Isomerases: Compact Versus Extended Active Sites in Ketosteroid Isomerase and Phosphoglucose Isomerase. *Biochemistry* 50, 9283–9295.
- (70) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- (71) Raines, R. T., Sutton, E. L., Straus, D. R., Gilbert, W., and Knowles, J. R. (1986) Reaction energetics of a mutant triose phosphate isomerase in which the active-site glutamate has been changed to aspartate. *Biochemistry* 25, 7142–7154.
- (72) Blacklow, S. C., and Knowles, J. R. (1990) How can a catalytic lesion be offset? The energetics of two pseudorevertant triosephosphate isomerases. *Biochemistry* 29, 4099–4108.
- (73) Blacklow, S. C., Liu, K. D., and Knowles, J. R. (1991) Stepwise improvements in catalytic effectiveness: Independence and interdependence in combinations of point mutations of a sluggish triosephosphate isomerase. *Biochemistry* 30, 8470–8476.
- (74) Cho, H.-S., Choi, G., Choi, K. Y., and Oh, B.-H. (1998) Crystal Structure and Enzyme Mechanism of  $\Delta^5$ -3-Ketosteroid Isomerase from *Pseudomonas testosteroni*. *Biochemistry* 37, 8325–8330.
- (75) Carugo, O., and Argos, P. (1997) Correlation between Side Chain Mobility and Conformation in Protein Structures. *Protein Eng.* 10, 777–787.
- (76) McLaughlin, R. N., Jr., Poelwijk, F. J., Raman, A., Gosai, W. S., and Ranganathan, R. (2012) The spatial architecture of protein function and adaptation. *Nature* 491, 138–142.
- (77) Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011) Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* 147, 1564–1575.
- (78) Rossman, M. G., Malcolm, A. D. B., and Bodmer, W. F. (1981) Evolution of Glycolytic Enzymes [and Discussion]. *Philos. Trans. R. Soc. B* 293, 191–203.
- (79) Lesk, A. M. (1995) NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* 5, 775–783.