ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/271590315

Prediction of distal residue participation in enzyme catalysis: Distal Residue Participation in Enzyme Catalysis

ARTICLE *in* PROTEIN SCIENCE · JANUARY 2015

Impact Factor: 2.85 · DOI: 10.1002/pro.2648

CITATIONS
3

READS
31

8 AUTHORS, INCLUDING:

Heather Brodkin
Alkermes
7 PUBLICATIONS **40** CITATIONS

SEE PROFILE

Srinivas Somarowthu
Yale University
13 PUBLICATIONS **97** CITATIONS

SEE PROFILE

Dagmar Ringe
Brandeis University
270 PUBLICATIONS **13,368** CITATIONS

SEE PROFILE

Mary Jo Ondrechen
Northeastern University
88 PUBLICATIONS **2,131** CITATIONS

SEE PROFILE

Available from: Walter RP Novak
Retrieved on: 26 January 2016

THE PROTEIN SOCIETY

Prediction of distal residue participation in enzyme catalysis

Heather R. Brodkin,[1,2,3†] Nicholas A. DeLateur,[1‡] Srinivas Somarowthu,[1§] Caitlyn L. Mills,[1] Walter R. Novak,[2,3¶] Penny J. Beuning,[1] Dagmar Ringe,[2,3] and Mary Jo Ondrechen[1]*

[1]Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115

[2]Department of Biochemistry, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454–9110

[3]Department of Chemistry, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454–9110

Received 22 November 2014; Accepted 26 January 2015
DOI: 10.1002/pro.2648
Published online 28 January 2015 proteinscience.org

Abstract: A scoring method for the prediction of catalytically important residues in enzyme structures is presented and used to examine the participation of distal residues in enzyme catalysis. Scores are based on the Partial Order Optimum Likelihood (POOL) machine learning method, using computed electrostatic properties, surface geometric features, and information obtained from the phylogenetic tree as input features. Predictions of distal residue participation in catalysis are compared with experimental kinetics data from the literature on variants of the featured enzymes; some additional kinetics measurements are reported for variants of *Pseudomonas putida* nitrile hydratase (ppNH) and for *Escherichia coli* alkaline phosphatase (AP). The multilayer active sites of *P. putida* nitrile hydratase and of human phosphoglucose isomerase are predicted by the POOL log *ZP* scores, as is the single-layer active site of *P. putida* ketosteroid isomerase. The log *ZP* score cutoff utilized here results in over-prediction of distal residue involvement in *E. coli* alkaline phosphatase. While fewer experimental data points are available for *P. putida* mandelate racemase and for human carbonic anhydrase II, the POOL log *ZP* scores properly predict the previously reported participation of distal residues.

Keywords: remote residues; catalytic efficiency; partial order optimum likelihood; nitrile hydratase; phosphoglucose isomerase; ketosteroid isomerase; alkaline phosphatase

Introduction

Many chemical reactions that require high temperature and/or extreme pH in the laboratory are catalyzed by enzymes at physiological temperature and neutral pH. One of the central challenges in biochemistry is to understand how enzymes achieve

†*Present address:* Potenza Therapeutics, 1030 Massachusetts Avenue, Suite 210, Cambridge, MA, 02138
‡*Present address:* Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139
§*Present address:* Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, 06520
¶*Present address:* Department of Chemistry, Wabash College, Crawfordsville, IN, 47933

Abbreviations: CSA, catalytic site atlas; INTREPID, information theoretic tree traversal for protein functional site identification; KSI, ketosteroid isomerase; PGI, phosphoglucose isomerase; POOL, partial order optimum likelihood; THEMATICS, theoretical microscopic anomalous titration curve shapes.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Grant sponsor: National Science Foundation; Grant number: MCB-1158176.

*Correspondence to: Mary Jo Ondrechen, Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115. E-mail: mjo@neu.edu

 Published by Wiley-Blackwell. © The Protein Society

this catalytic power under mild conditions. Structural and biochemical studies have now characterized the active sites of hundreds of enzymes; nearly all of these studies have focused on the amino acid residues in direct contact with the reacting substrate molecule(s). A typical enzyme might consist of hundreds of residues, but only a handful of them have been identified in the literature as catalytically important. In the manually curated portion of the Catalytic Site Atlas (CSA),[1] a database of enzyme active site composition compiled from the experimental literature, an average of only 1.2% of the residues in the biological units of these 170 enzymes is listed as catalytically important. Virtually all of these reported catalytic residues are, in the active conformation of the enzyme, in direct contact with the reacting substrate molecule(s). A recent review surveys examples from the literature of proteins where distal residues have been shown to participate directly in the biochemical function.[2] This work presents a model, based on computed chemical properties, that successfully predicts both first-shell and more distant residues that are important for function. Model predictions are compared with available experimental data and, for a few cases, new experimental results are presented. The ability to predict, with a simple calculation, the involvement of distal residues in catalytic and other processes is a very important capability for protein engineering and for understanding how nature builds catalytic sites.

Here new computational and experimental results, together with experimental data from the literature, are presented to show that residues that are not in direct contact with the reacting substrate molecule often play significant roles in catalysis, but more importantly, that their participation in catalysis is predictable with a simple calculation. Previously we have shown[3,4] that all of the residues in a protein structure may be assigned scores based on computed chemical properties and that these scores may be used to rank-order all of the residues according to their probability of functional importance. In the present work, these scores are transformed so that they are more comparable across proteins of different sizes. We now show that the pattern of these scores predicts whether an enzyme has a highly localized, compact active site, or a spatially extended, multilayer active site with extensive involvement of distal residues in catalysis and ligand binding.

For purposes of the present study, residues interacting directly with the substrate molecule are referred to as first-shell residues; other residues interacting directly with one or more first-shell residues are called second-shell residues; and finally, other residues interacting with one or more second-shell residues are called third-shell residues. All residues defined as first-, second-, or third-shell thus may be thought of as members of respective spheres surrounding the reacting substrate molecule. In the present work, enzymes with available experimental data on the participation of distal residues in catalysis are the focus.

THEoretical Microscopic Anomalous TItration Curve Shapes (THEMATICS) is a computational method for the prediction of catalytic and binding residues in proteins solely from the three-dimensional structure of the query protein.[5–7] THEMATICS takes advantage of the unique chemical and electrostatic properties of catalytic and binding residues. In particular, the ionizable residues in active sites tend to have anomalously shaped theoretical titration curves. We have established metrics[6] to quantify the degree of deviation from normal titration behavior and have shown how these metrics may be used to predict, with high sensitivity and specificity,[7] the active site residues in a protein three-dimensional structure, just from a simple set of calculations.

Partial Order Optimum Likelihood (POOL)[3,4,8] is a machine learning method for the prediction of functionally important residues in protein structures. POOL, a multidimensional, monotonicity-constrained maximum likelihood technique, utilizes input for which the output variable (in this case the probability of functional importance of a residue) is a monotonic function of each of the input features. POOL starts with the hypothesis that the larger the THEMATICS metrics for a given residue, the higher the probability that that residue is important for the biochemical function. Electrostatics features from THEMATICS are combined with multidimensional isotonic regression to form maximum likelihood estimates of probabilities that specific residues are active in biochemical function. All residues are then rank-ordered according to their probability of functional importance. Through the introduction of environment variables in POOL, THEMATICS metrics are used to characterize the chemical environments of all residues, not just the ionizable ones. Furthermore, the input to POOL is flexible, in that any property upon which the probability of the functional importance of a residue depends monotonically can be a POOL input feature. These include features based solely on the structure of the query protein, such as THEMATICS metrics[6,7,9] and surface geometric properties,[10] but may also include other features if they are available, such as phylogenetic tree-based scores.[11] It has been demonstrated[3] that POOL with purely structure-based features performs very nearly as well as the best methods that utilize both sequence alignments and structure. POOL with input features that include THEMATICS, geometric data, and phylogenetic scores has been shown[4] to be an even higher-performing functional residue predictor; with these three input

features, THEMATICS is the largest contributor to the POOL predictions. In this work, THEMATICS, the structure-only version of ConCavity,[10] and INTREPID,[11] are used as input to POOL to predict specific residues, including those outside the first shell, that are involved in catalysis; POOL scores are used to predict the degree of spatial extension of enzyme active sites and experimental evidence to support these predictions is presented.

Numerous examples of distal residue involvement in enzyme specificity have been reported in the directed evolution literature.[12] For instance the activity of human carbonic anhydrase II on the ester substrate 2-naphthyl acetate was increased 40-fold with a triple mutant obtained through three rounds of mutagenesis, selection and recombination. Notably, all three residues were located outside of the first shell.[13] In the directed evolution of *Escherichia coli* D-sialic acid aldolase to L−3-deoxy-*manno*−2-octulosonic acid aldolase, changes in eight amino acids, all of which are located outside the first shell, were necessary to produce a mutant enzyme with a 1000-fold increase in specificity for the unnatural sugar substrate, L–D−3-deoxy-*manno*−2-octulosonic acid.[14] Directed evolution techniques were also used to alter the specificity of aspartate aminotransferase.[15] An enzyme variant containing 17 amino acid substitutions resulted in a 10^6-fold increase in catalytic rate for the non-native valine substrate.[15] Interestingly, only one of the mutated residues interacted directly with the substrate; all others were remote residues. Random mutagenesis experiments on the flavoenzyme vanillyl-alcohol oxidase generated four single point mutants, each with enhanced reactivity for creosol and other *ortho*-substituted 4-methylphenols.[16] In each mutant enzyme, the mutation was located outside of the first shell. Finally, a mutant metallo-β-lactamase was discovered through directed evolution that resulted in an enzyme with increased hydrolytic efficiency toward cephalexin.[17] This mutant contained four amino acid substitutions, two in the second coordination sphere of the metal ion and two far removed from the annotated active site.

There are also a few examples where site-directed mutagenesis studies have shown that distal residues contribute to reactivity. In the metalloenzymes alkaline phosphatase[18–21] and mandelate racemase,[22] rational mutations to residues in the second shell surrounding the metal ions have been reported previously, sometimes with dramatic results. Studies showing participation in catalysis by multiple THEMATICS-predicted second- and third-shell residues have been reported recently by us for *Pseudomonas putida* nitrile hydratase[23] and human phosphoglucose isomerase,[24] and also for the Y-family DNA polymerase from *E. coli* DinB.[25] Further analysis of these and additional enzymes are reported here, using the new POOL scores, new experimental data, and compilations of experimental data from the literature.

Results

Optimizing POOL scores to predict distal residues

The top 8% of POOL-ranked residues typically have been taken to be the predicted set of functionally important residues because this gives the best performance against the CSA-100,[11] the benchmark set of annotated catalytic residues, for purposes of predicting the known catalytic residues. The CSA-100 is a 100-enzyme subset of the original, manually curated portion of the Catalytic Site Atlas;[1] the 100 enzymes were chosen to maximize sequence diversity with the goal of no detectable homology between any two members of the subset. However, virtually all of the CSA-100 annotated residues are first-shell residues and thus the cutoff that gives optimum performance on the benchmark set is not necessarily the best cutoff for the prediction of all of the residues that participate in catalysis, including the distal residues. For purposes of prediction of both first-shell and more distant residues, we introduce the POOL scores, which contain useful information and enable better predictions of functionally important residues, both inside and outside the first shell. Note that POOL scores are used to rank-order the residues within a given protein structure, according to their probability of participation in catalysis. However, the raw POOL scores are not comparable between two different proteins, as they depend on protein size. It is desirable to transform the POOL scores to a scale that is comparable across proteins. Therefore the POOL scores within each protein are linearly transformed into Z scores. Generally the few highest-ranked residues have Z scores that are many orders of magnitude larger than those of most of the other residues. Therefore, the mean and standard deviation σ_P of all of the POOL scores within a protein are calculated a first time and the residues with POOL scores greater than two standard deviations over the mean are deemed outliers. The trimmed set of POOL scores, excluding the outlier residues, is then used for the second calculation of the mean and standard deviation. The POOL scores for all residues are then expressed as Z scores, as:

$$Z_k = (P_k - \langle P \rangle_t)/\sigma_{Pt}, \quad (1)$$

Here k is a residue index, $_t$ represents the mean POOL score calculated on the trimmed set and σ_{Pt} represents the standard deviation for the trimmed set. Because the range of POOL scores is over many orders of magnitude in a given protein, the Z scores are re-written as a logarithmic function log ZP as:

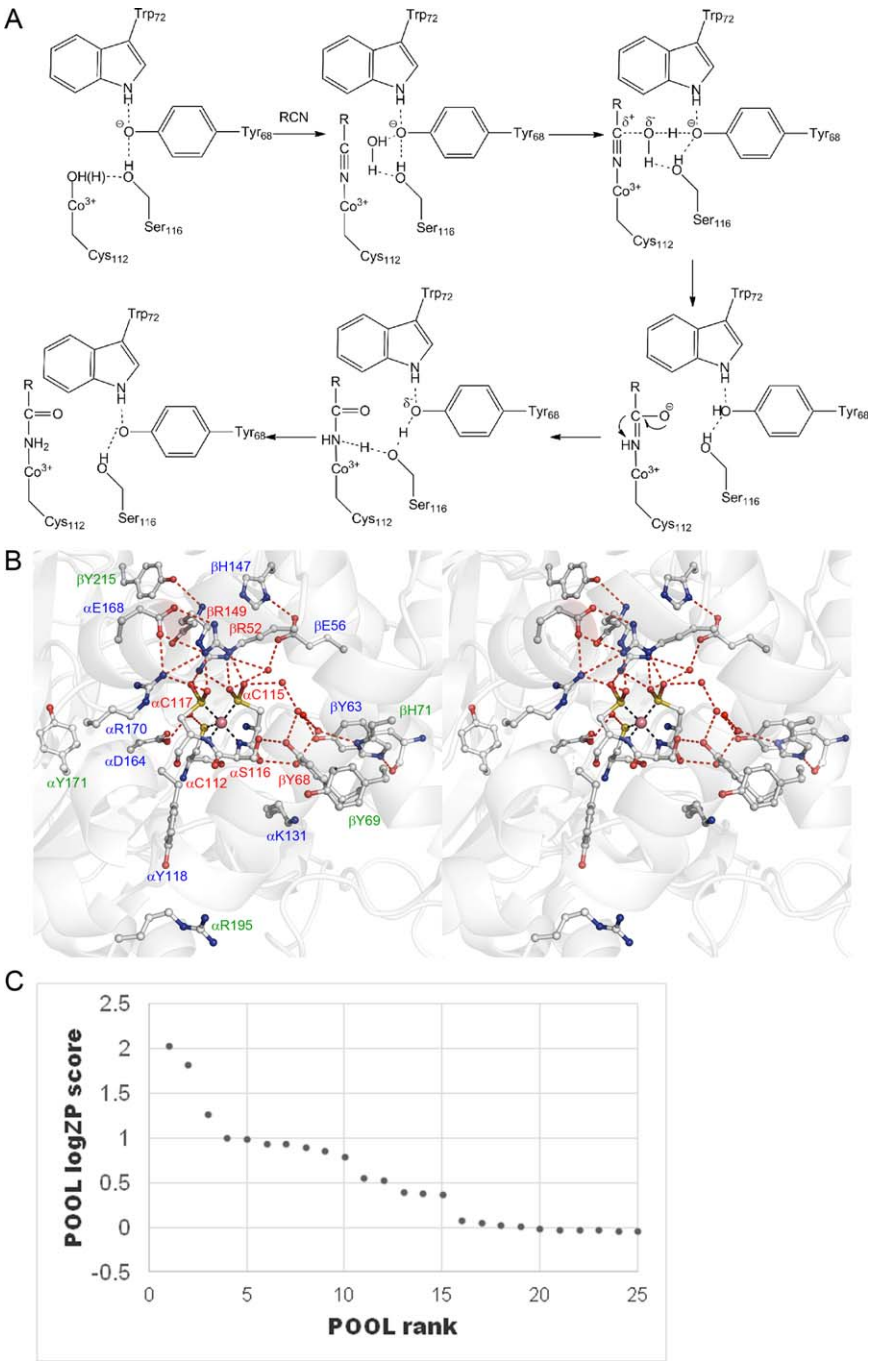


Figure 1. For *Pseudomonas putida* Nitrile hydratase: A) reaction scheme; B) stereo view of the extended active site. The Co ion is rendered as a pink ball; font color indicates location—red, first sphere—blue, second sphere—green, third sphere; C) POOL log *ZP* score as a function of POOL rank for the top 25 POOL ranked residues.

$$\log ZP_k = \log_{10}(Z_k + 1) \quad (2)$$

Using the new POOL scores, we analyze enzymes across different enzyme classes.

Nitrile hydratase

Nitrile hydratase (E.C. 4.2.1.84, NHase) from *P. putida* utilizes low-spin, noncorrinoid Co(III) to catalyze the hydrolysis of nitrile substrates to their corresponding amides. The reaction scheme is shown in Figure 1(A). This type of low-spin Co(III) NHase is prevalent as a biocatalyst for the industrial production of commodity chemicals such as acrylamide on the kiloton scale, and also nicotinamide and 5-cyanovaleramide.[26] Catalysis by NHase affords a number of advantages over conventional catalysts, including low energy consumption, less waste, high efficiency and product purity. *P. putida* NHase (ppNH) is a heterodimeric metalloenzyme. The two subunits, α and β, do not show mutual homology but both α and

Table I. Residue POOL Rank, POOL log ZP Score, and Loss of Catalytic Efficiency for ppNH Variants

Variant	Shell	POOL rank	POOL log *ZP* score	$(k_{cat}/K_M)_{wild\ type}/(k_{cat}/K_M)_{mutant}$	Reference
α-C115A	First	1	2.0	$>1.0 \times 10^5$	27
β-Y68F	First	2	1.8	3.8×10^2	28
α-C112A	First	3	1.3	$>1.0 \times 10^5$	27
β-R149Y	First	4	1.0	See text	30; See text
β-Y69F	Third	5	0.99	6	This work
α-K131Q	Second	6	0.94	11	This work
β-Y63F	Second	7	0.94	11	This work
β-E56Q	Second	8	0.89	2.3×10^2	23
β-H147N	Second	9	0.85	1.1×10^2	23
α-C117A	First	10	0.79	$>1.0 \times 10^5$	27
α-D164N	Second	11	0.56	20	23
β-R52K	First	12	0.52	1.3×10^3	29
β-H71L	Third	15	0.37	25	23
α-Y118F	Second	17	0.047	8	This work
α-E168Q	Second	22	−0.029	3.7	23
β-Y215F	Third	32	−0.054	3.3	23
α-R170N	Second	35	−0.062	1.6	23
α-Y171F	Third	40	−0.067	2.1	23
α-R195Q	Third	51	−0.072	3	This work

Note that the residue ranked fourth, β-R149, has been reported to be important for catalysis, but numerical kinetics data have not been reported.

β do show a high degree of homology to other known NHases.

Six first-shell residues in ppNH were reported previously[27–30] to be important for catalysis: the cobalt-coordinating α-C112, α-C115, and α-C117, the catalytic β-R52 and β-R149, and the ligand-binding β-Y68. The two arginine residues are thought to hydrogen bond to the cysteine residues which coordinate the metal ion. These arginine residues therefore appear to stabilize the claw setting in the active site. We previously reported[23] first-, second-, and third-shell residues predicted by THEMATICS[7] to be important for catalysis in ppNH (PDB ID: 3QXE) through site-directed mutagenesis (SDM) experiments. A total of eight residues were tested, five of which were predicted by THEMATICS and three were not predicted; all eight residues are well conserved. Mutations, made as conservatively as possible, at four of the five THEMATICS-predicted residues showed significant loss of catalytic activity; one predicted residue, plus the three residues not predicted, did not show significant difference from wild type upon mutation. In the present work, we report additional experiments on distal residues, using the POOL scores as a guide. The effects of the mutations are summarized in Table I. A stereo view of the extended active site of ppNH is shown in Figure 1(B). The cobalt ion is rendered as a pink ball. The first-sphere residues are labeled in red font, the second-sphere in blue, and the third sphere in green.

Figure 1(C) shows the POOL log *ZP* scores as a function of POOL rank for ppNH. The top 19 residues, corresponding to the top 5% of all residues in the dimer, have positive log *ZP* scores. For ppNH, 10 out of the 15 top POOL-ranked residues were previously tested by site-directed mutagenesis, either reported in the prior literature or by us, and all 10 of them show a significant effect on catalysis. These 10 include the 6 first-shell residues reported previously (the cobalt-coordinating α-C112, α-C115, and α-C117, the catalytic β-R52 and β-R149, and the ligand-binding β-Y68).[27–30] Note that in Fe-type nitrile hydratase, the residue equivalent to β-R149 of the Co-type nitrile hydratase analyzed here, β-R141, has been reported[30] to show no activity upon mutation to Y or E, and significant decrease in activity upon mutation to K, although no specific kinetics values were provided. The remaining four residues corresponding to significant loss of activity upon mutation include three second-shell residues (α-D164, β-E56, and β-H147) and one third-shell residue (β-H71) reported by us to show a loss in catalytic efficiency of at least 20-fold.[23] α-D164 interacts with α-C112 through a water molecule; mutation to N retains the relative size, shape and hydrogen bonding capability while removing the ability to form a negative charge. β-E56 is within hydrogen bonding distance to α-C115, β-R149 and β-H147, while the second-shell residue β-H147 is located behind both catalytic residues β-R52 and β-R149. The third-shell residue β-H71 interacts through water molecules with the metal coordinating residue, α-C115. Mutation of β-H71 to L removes hydrogen bond capabilities. In contrast, there are four variants reported by us which exhibit an insignificant change in catalytic efficiency: second-shell residues α-E168 (involved in a salt bridge to β-R52) and α-R170 (H-bonded to α-C117); and third-shell residues β-Y215 (H-bonded to the side chain of β-R149

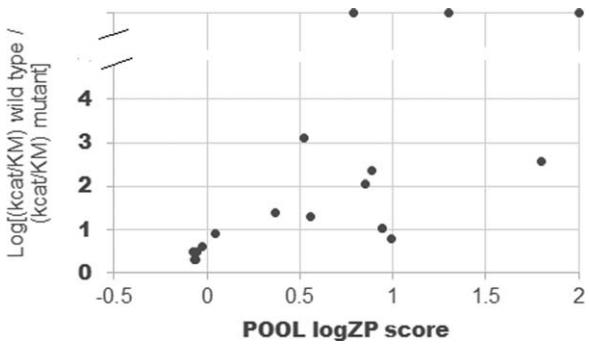


Figure 2. Log_{10} of the ratio of the catalytic efficiency of WT over variant as a function of the POOL log *ZP* score of the substituted position in the variant for ppNH.

and to distal residue β-D172) and α-Y171 (H-bonded to distal residue β-D210). These four residues are all ranked lower by POOL (ranking 22nd, 32nd, 35th, and 40th, respectively), with negative log *ZP* scores.

Some additional residues with high POOL ranks, β-Y69, α-K131, and β-Y63, ranked fifth, sixth, and seventh, respectively, have now been studied by SDM. α-K131 is a second-shell residue located behind the first-shell residue β-Y68 with respect to the substrate. β-Y69 is a third-shell residue 11.9 Å from the Co ion and located behind α-K131 with respect to the substrate. β-Y63 is a second-shell residue behind the Co-coordinating α-C115. Results of these mutations are summarized, together with prior results, in Table I and illustrated in Figure 2. Figure 2 plots log_{10} of the ratio of the catalytic efficiency for wild-type ppNH to that of the variant, as a function of the POOL log *ZP* score of the residue at the substituted position in the variant. The second-shell variants α-K131Q and β-Y63F both show an 11-fold reduction in catalytic efficiency. The third-shell variant β-Y69F shows a small but significant loss in catalytic efficiency.

Notice in Figure 2 that the effect on catalytic efficiency generally declines with declining POOL score and falls off to a kinetically small or insignificant effect for negative values of the POOL log *ZP* score. (A threefold or less decrease in catalytic efficiency is not considered significant.) It is important to note in Table I that there are second- and third-shell residues with positive POOL log *ZP* scores that do show a significant effect on catalytic efficiency. Thus the POOL log *ZP* scores give some guidance about the number of residues that participate significantly in catalysis; the residues with the higher, positive scores tend to show the most participation in catalysis whereas residues with negative scores tend not to show significant effects.

Phosphoglucose isomerase

Phosphoglucose isomerase (PGI; E.C. 5.3.1.9) is an example of a moonlighting protein[31] that has multiple functions. Inside the cell, PGI catalyzes the reversible isomerization of glucose-6-phosphate to fructose-6-phosphate [Fig. 3(A)], an essential step in glycolysis. Extracellular PGI has many roles including neuroleukin,[32,33] autocrine motility factor,[34] and maturation factor.[35] Human phosphoglucose isomerase (hPGI) is medically important because certain mutations in the gene that encodes for hPGI have been identified to cause nonspherocytic hemolytic anemia.[36] In addition, hPGI is used as a biomarker because of its role as a tumor-secreted cytokine and angiogenic factor in certain types of cancer.[37] The structure of hPGI is a dimer of two identical subunits.[38]

Figure 3(B) shows a stereo view of the extended active site of human PGI (hPGI) and Figure 3(C) shows the POOL log *ZP* scores as a function of POOL rank for hPGI. Like ppNH, the POOL log *ZP* scores for hPGI exhibit a sharp fall-off after the first few residues, but then an elongated tail, indicating a prediction of an active site with many participating residues, including residues outside the first layer. The top 47 residues, corresponding to the top 8%, have positive log *ZP* scores.

Experiments confirmed the prediction of a spatially extended active site, as shown in Table II. The catalytic residues for PGI were originally established by site-directed mutagenesis experiments[39,40] on the *Bacillus stearothermophilus* enzyme. Catalytically active residues for mammalian PGI were then identified by sequence alignment and by X-ray structure determination.[38,41–43] Although the human and *Bacillus stearothermophilus* enzymes share only 21% sequence identity, their active site residues are well conserved. Kinetics data in Table II are for human PGI, except where noted. The seven cases in parentheses (R273, E358, E217, K519, Q512, K211, and K440) are for the homologous residues in the *B. stearothermophilus* enzyme (R207, E290, E150, K425, Q418, K144, and K356, respectively) and all except for K440 are presumed to be important also in human PGI. Figure 4 shows a plot of log_{10} of the ratio of the catalytic efficiency for wild-type PGI to that of the variant, as a function of the POOL log *ZP* score of the residue at the substituted position in the variant.

For hPGI, important first-shell residues, H389, R273, E358, E217, K519, Q512, and K211, all rank among the top 45 residues and all have positive POOL log *ZP* scores. Second-shell residues, E495, K362, and Q388, and third shell residues, H396 and H100, that have been shown to be important for catalysis,[24] are also among the highest ranking residues. The second-shell residue D511, that has been shown[24] to play a significant role in catalysis, is ranked 57th and falls just below the range of residues with positive scores. There are three highly ranked second- and third-shell residues, Y341, Y274,

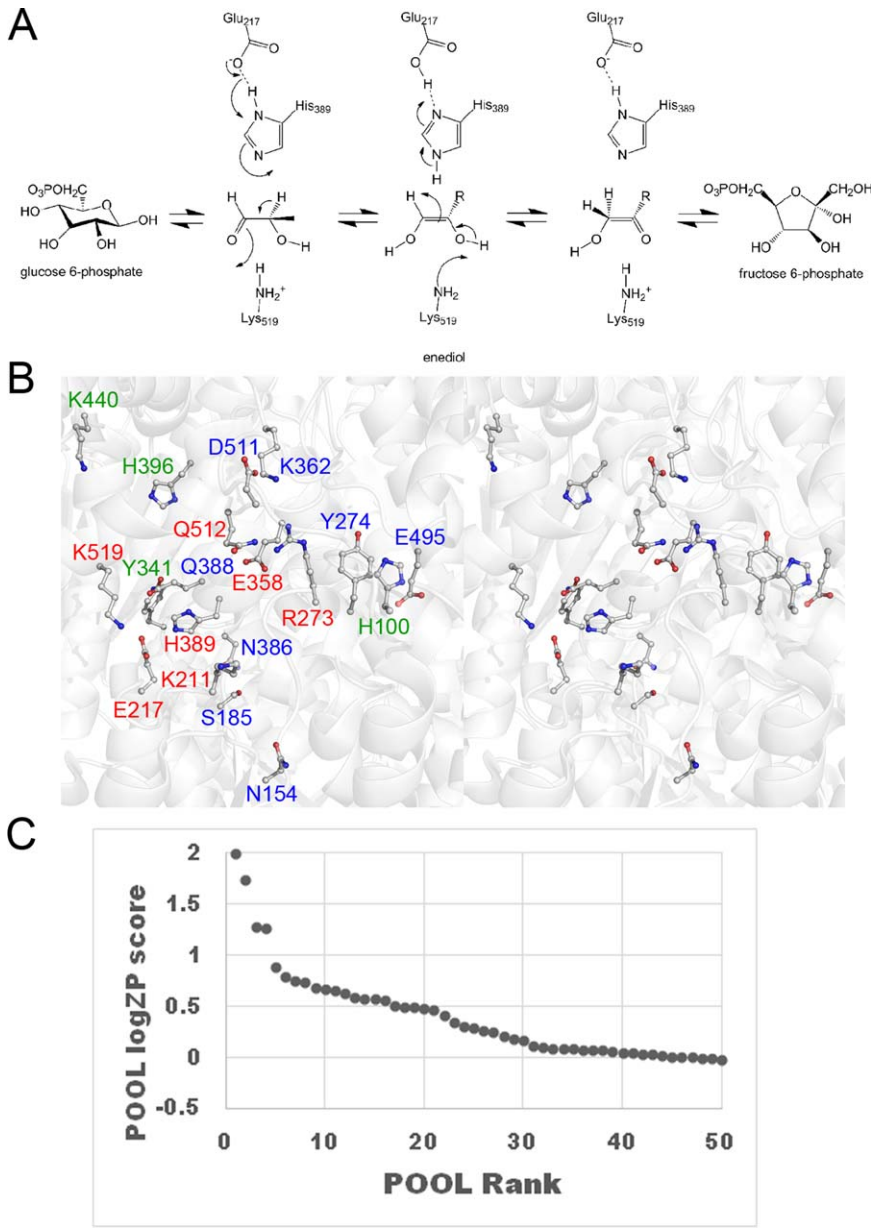


Figure 3. For human Phosphoglucose isomerase: A) reaction scheme; B) stereo view of the extended active site—font color indicates location—red, first sphere—blue, second sphere—green, third sphere; C) POOL log *ZP* score as a function of POOL rank for the 50 top-ranked residues.

and N386 that proved to be false positives, at least with respect to single-site mutation. Residues that are highly conserved but ranked well below those with positive scores, N154, S185, and K440, are true negatives and the effects of mutation on catalysis are not significant.

Ketosteroid isomerase

Δ^5−3-Keto steroid isomerase (KSI; EC 5.3.3.1) is a well-studied[44–51] enzyme with high catalytic activity, with a rate approaching the diffusion-controlled limit. KSI catalyzes the isomerization of a Δ^5−3-ketosteroid to a Δ^4−3-ketosteroid [Fig. 5(A)], involving C-H bond cleavage and formation through a dienolate intermediate. KSI has a metabolic role in the degradation of steroids in bacterial species that can live on steroids as their sole carbon source. High-resolution crystal structures of KSI from *P. putida* (ppKSI, PDB ID: 1OPY[52]) and *Commamonas testosteroni* (PDB ID: 1QJG[53]) reveal that, although these two enzymes share only 34% sequence identity, their active site residues are conserved and structurally superimposable. Figure 5(B) shows a stereo view of the active site of ppKSI.

While ppKSI and hPGI both catalyze isomerization reactions and are both members of EC class 5.3, these two isomerases possess very different types of active sites.[24] In the plot of POOL log *ZP* score as a function of POOL rank for the residues of ppKSI,

Table II. Residue POOL Rankings, POOL log ZP Score, and Loss of Catalytic Efficiency for PGI Variants

Variant	Shell	POOL rank	POOL log *ZP* score	$(k_{cat}/K_M)_{wild\ type}/(k_{cat}/K_M)_{mutant}$	Reference
H389L	First	1	2.0	>10,000	24
H396L	Third	3	1.3	24	24
(R273A)	First	6	0.79	(170,000)	39
Y341F	Third	11	0.66	1.5	24
(E358Q)	First	13	0.59	(2,100)	40
E495Q	Second	14	0.58	250	24
(E217Q)	First	20	0.49	(130,000)	40
H100L	Third	21	0.48	630	24
Y274F	Second	23	0.34	2.2	24
N386A	Second	25	0.22	2.8	24
K362A	Second	26	0.27	>10,000	24
Q388A	Second	29	0.18	19	24
(K519A)	First	39	0.06	(390)	39
(Q512A)	First	44	0.02	(120)	40
(K211A)	First	45	0.01	(19)	39
D511N	Second	57	−0.03	230	24
N154Q	Second	135	−0.08	3.8	24
S185A	Second	140	−0.08	3.5	24
(K440A)	Third	162	−0.08	(1.0)	39

Parentheses indicate kinetics data reported for the homologous residue in the *B. stearothermophilus* enzyme (with hPGI numbering). All other kinetics data are for human PGI.

shown in Figure 5(C), the residues of ppKSI exhibit a sharp fall-off after the first few residues and no extended tail, corresponding to a prediction of a compact active site. The 13 highest scoring residues, comprising 10% of this small protein, have positive log *ZP* scores. The top three residues in the POOL ranking, Y32, Y57, and Y16, form an important hydrogen bonding network in the active site of KSI;[50] Y16 stabilizes the dienolate intermediate and serves as a proton donor. The Y16F variant shows a 550-fold reduction in k_{cat}/K_M relative to wild type;[54] the Y16S variant shows an 18-fold reduction.[55] While the Y32F and Y57F variants do not show major differences from wild type, with 1.3- and 2.9-fold reductions[54] in k_{cat}/K_M, respectively, k_{cat}/K_M for the Y57S variant is 66-fold less[55] than wild type (see Table III). Table III and Figure 6 show that mostly first-shell residues participate in catalysis; residues outside the first shell have only small effects, as predicted.

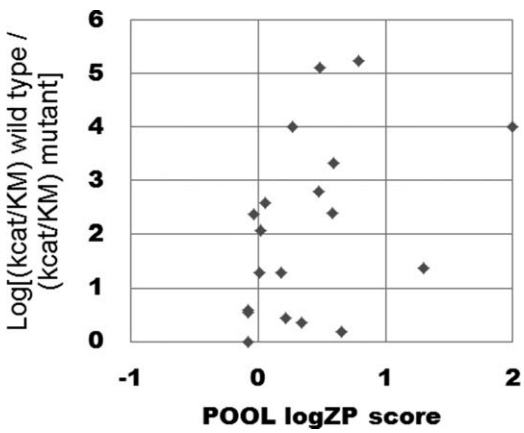


Figure 4. Log 10 of the ratio of the catalytic efficiency of WT over variant as a function of the POOL log *ZP* score of the substituted position in the variant for human phosphoglucose isomerase.

Alkaline phosphatase

Alkaline phosphatase (AP, EC 3.1.3.1) has been widely studied using both directed evolution and rational protein design approaches.[21] AP is a dimeric metalloenzyme containing two zinc ions and one magnesium ion per subunit and it catalyzes the reversible hydrolysis of phosphomonoesters to yield inorganic phosphate plus an alcohol; the reaction mechanism is shown in Figure 7(A). In the crystal structure of *E. coli* AP (PDB ID: 1ALK[59]), there is an inorganic phosphate ion bound to the two zinc ions (Zn_1 and Zn_2) and to the guanidinium group of R166.[19] The active site region of the protein includes D101, S102, A103, the three metal ions and their coordinating residues, and R166;[59] R166 has been shown to be important for transition state stabilization and for discrimination between phosphates and sulfates.[60,61] There is a hydrogen bond network that includes hydrogen bonds between D101 and R166 and between R166 and a water molecule.[62] K328 interacts with the phosphate group in the active center through a water mediated hydrogen bond. A stereo image of the active site is shown in Figure 7(B).

Two interesting mutations, designed specifically to increase the catalytic activity of AP, are associated with the first-shell residue D153 and the second-shell residue D330. D153 is involved in an ion pair interaction with K328 and in a water-mediated interaction with the Mg^{2+} ion.[21] D153 also

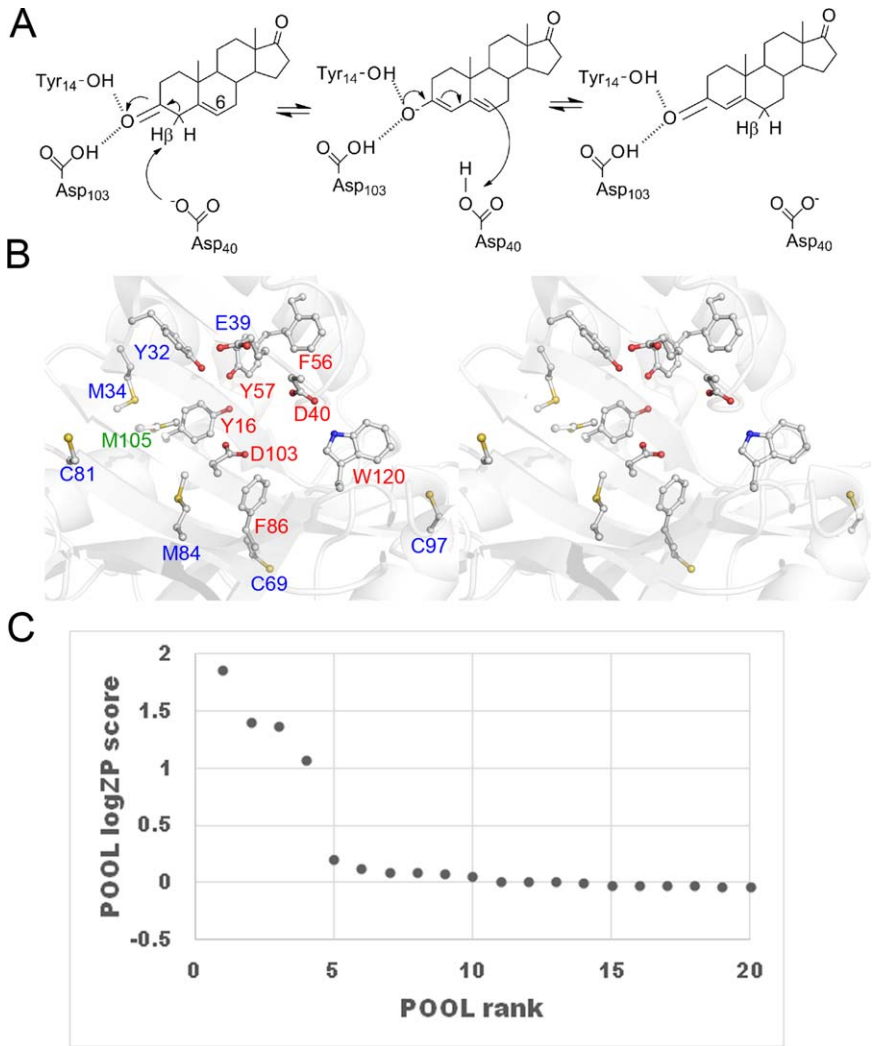


Figure 5. For *Pseudomonas putida* Ketosteroid isomerase: A) reaction scheme; B) stereo view of the extended active site—font color indicates location—red, first sphere—blue, second sphere—green, third sphere; C) POOL log *ZP* score as a function of POOL rank for the top 20 POOL-ranked residues.

Table III. Residue POOL Rankings, POOL log ZP Score, and Change in Catalytic Efficiency for *P. putida* KSI Variants

Variant	Shell	POOL rank	POOL log *ZP* score	$(k_{cat}/K_M)_{wild\ type}/(k_{cat}/K_M)_{mutant}$	Reference
Y32F	Second	1	1.9	1.3	54
Y32S	"	"	"	3.6	55
Y57F	First	2	1.4	2.9	54
Y57S	"	"	"	66	55
Y16F	First	3	1.4	550	54
Y16S	"	"	"	18	55
D40N	First	4	1.1	334,000	56
W120F	First	5	0.20	13	57
F56L	First	6	0.13	3.6	57
D103N	First	9	0.074	27	51
E39Q	Second	12	0.015	2.4	24
F86L	First	14	−0.006	9.2	57
M105A	Third	16	−0.026	7.6	24
M84A	Second	23	−0.05	0.98	24
M31A	Second	34	−0.05	2.1	24
C97S	Second	39	−0.05	1.3	58
C81S	Second	48	−0.05	0.98	58
C69S	Second	58	−0.05	0.92	58

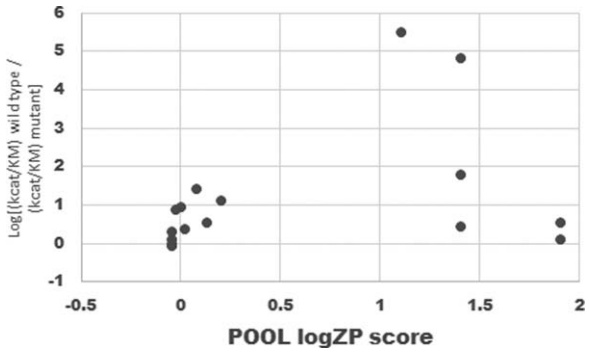


Figure 6. Log 10 of the ratio of the catalytic efficiency of WT over variant as a function of the POOL log *ZP* score of the substituted position in the variant for *P. putida* ketosteroid isomerase.

serves to position the catalytic residue, R166. Using rational protein design methods, four different single-point mutations were made at the D153 position in an attempt to understand more clearly its role and to increase the catalytic efficiency of the protein.[63–65] Three of the four mutations, D153G, D153E, and D153A, resulted in an increased catalytic rate.

In an effort to increase the turnover rate to a level comparable to that of mammalian enzymes, while allowing the protein to retain its high thermostability, Muller *et al.*[21] assumed that while mutations to residues in the active site most often had a negative effect on catalysis, residues outside the catalytic site may also influence catalysis. Using directed evolution approaches, they showed that variants with double or triple mutations of D153, K328, and D330 showed increased activity over wild type while still maintaining thermostability.[21]

Figure 7(C) shows the POOL log *ZP* scores as a function of POOL rank for *E. coli* AP. The top 36 residues, or 8% of all residues, have positive log *ZP* scores; these top 36 include the two catalytic residues S102 and R166, D101 in the active site, and the eight Zn- and Mg-coordinating residues D51, T155, E322, D327, H331, D369, H370, and H412. Note that the metal-coordinating residues H370 and E322 are ranked third and fifth, respectively, by POOL (kinetics data not available). Also included is the second-shell residue H372, which hydrogen bonds to the Zn-coordinating residue D327. Mutation of H372 to alanine shows no change in Zn binding affinity, but k_{cat} of the H372A variant is ninefold less than that of wild type, with a 3.7-fold increase in catalytic efficiency (k_{cat}/K_M).[18]

In this work, two additional second-shell residues in the top 36, H86 and M53, the third-shell residue E57 at rank 17, and the lower-ranked second-shell residues Q435, E150, and S105, were studied by site-directed mutagenesis. Additional variants were also made for H372 and D330, two second-shell residues previously shown to influence activity. Table IV shows the POOL rank, POOL log *ZP* score, and change in catalytic efficiency for some *E. coli* AP variants. In this case, the log $ZP = 0$ cutoff for POOL scores results in over-prediction of the essential residues for catalysis. While all of the residues known to have significant, single-handed effects on activity are included in the top 20 predicted by POOL, predicted residues H86, E57, and M53 proved to be false positives with respect to single-point mutations.

Other examples of the importance of distal residues

Mandelate racemase. Mandelate racemase (EC 5.1.2.2) catalyzes the interconversion of the (*R*) and (*S*) enantiomers of mandelic acid (Scheme 1) via abstraction of a proton from the α-carbon atom.[72–74] It is a divalent cation-dependent protein and in the *P. putida* crystal structure (PDB ID: 2MNR), the Mn^{2+} ion is coordinated by D195, E221, E222, and E247.[75] The reaction proceeds via a two-base mechanism; H297 acts as the (*R*)-specific acid/base catalyst[73] while K166 acts as the (*S*)-specific acid/base catalyst.[74] The second-shell residue D270 forms a hydrogen bond with the catalytic H297. The single mutation D270N results in a 10,000-fold decrease in enzyme activity compared with wild type for both (*R*)- and (*S*)-mandelate substrates.[22] The N270 side chain in the structure of the variant is superimposable on the D270 side chain in the structure of the wild type; the remainder of the two structures are identical except that the side chain of the catalytic H297 is "rotated and displaced toward the binding site" in the mutant structure.[22] It was argued that D270 is necessary to impart the correct pK_a to the catalytic H297.[22] The two catalytic bases H297 and K166 and the important second-shell residues D270 are all highly ranked by POOL, with log *ZP* scores of 0.71, 0.65, and 0.65, respectively. S1

Carbonic anhydrase isoform II. Carbonic anhydrase isoform II (CAII, EC 4.2.1.1) is a zinc-dependent metalloenzyme that catalyzes the reversible hydration of carbon dioxide (Scheme 2).[76] The hydrolysis of carbon dioxide to the bicarbonate ion and a proton occurs via a two-step mechanism. In the first step, a zinc-bound hydroxide ion acts as the nucleophile which attacks CO_2 to form a zinc-bound bicarbonate intermediate. This intermediate is then displaced by a water molecule creating a zinc-H_2O complex and free bicarbonate ion. In the rate determining step, the zinc-bound hydroxide is regenerated through the transfer of a proton to the solvent facilitated by the active site histidine residue, H64, which acts as a proton shuttle.[77,78] In the crystal structure of CAII, (PDB ID: 1CA2[79]) the zinc ion is tetrahedrally coordinated with the side chains of S2

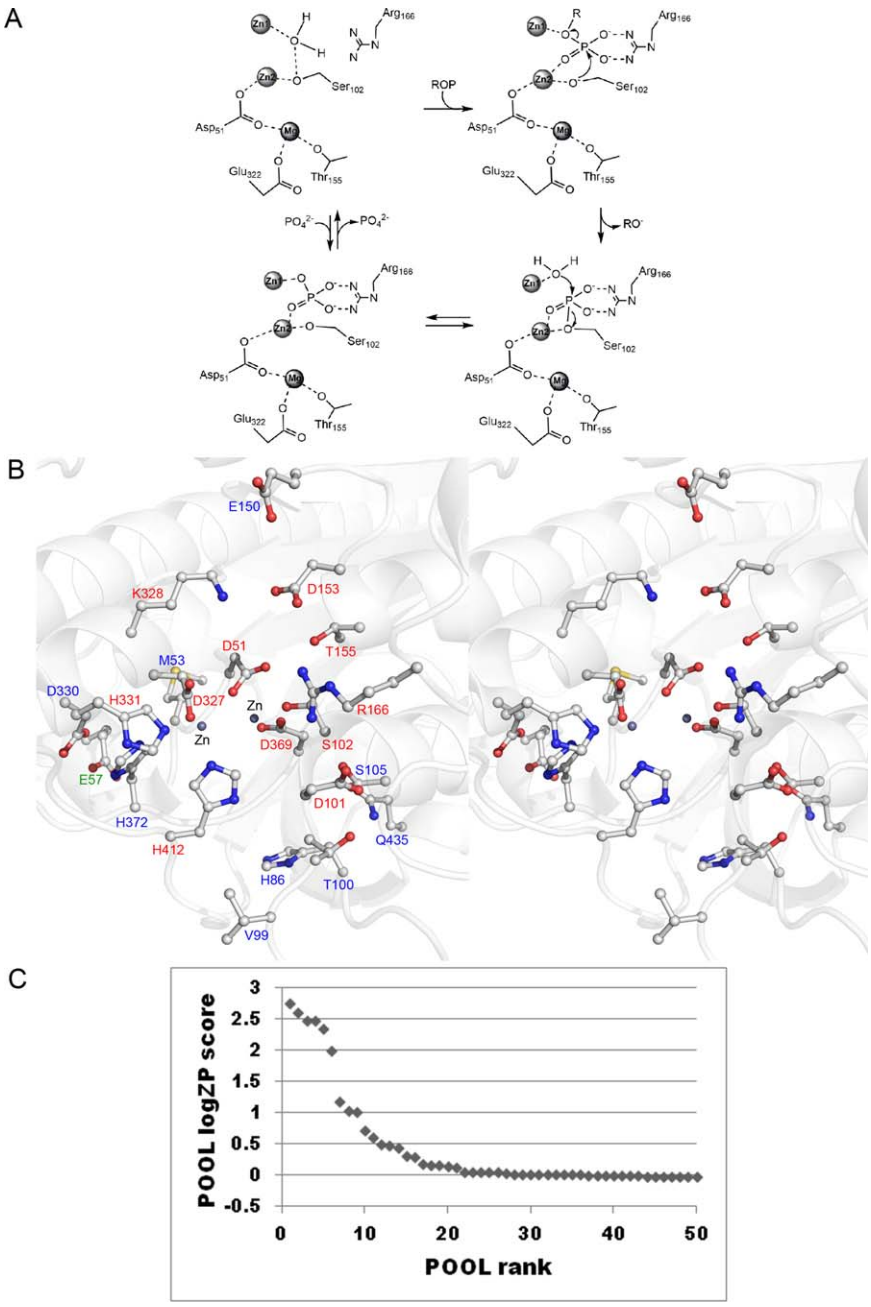


Figure 7. For *E. coli* alkaline phosphatase: A) reaction scheme; B) stereo view of the active site. The Zn ions are rendered as grey balls; Font color indicates location—red, first sphere—blue, second sphere—green, third sphere; C) POOL log *ZP* score as a function of POOL rank for the top 50 POOL-ranked residues.

H94, H96, and H119 and a hydroxide ion. T199 accepts a hydrogen bond from the zinc-bound hydroxide ion and donates a hydrogen bond to the side chain of E106. T199 also helps to orient the hydroxide ion properly for nucleophilic attack on the substrate, CO_2.[80] The second-shell residue H107 when mutated to Y causes CAII deficiency syndrome in vivo.[81] For another second-shell residue E106, loss-of-charge variants E106Q and E106A show decreased catalytic rate constants by factors of 850 and 110, respectively; the variant E106D that maintains the negative charge shows no change in catalytic rate from that of wild type.[82]

For CAII, only THEMATICS and ConCavity data were used in the POOL analysis because of residue numbering disparities in the INTREPID data. The top 19 residues have positive log *ZP* scores. The Zn-coordinating histidine residues H94, H119, and H96 are the top three ranking residues in POOL. The important second-shell residue E106 ranks fourth with a log *ZP* score of 1.9. The proton shuttle H64 ranks sixth with a log *ZP* score of 0.91. H107 in the second shell, mutation of which is associated with disease, ranks seventh with a log *ZP* score of 0.58. The hydroxide activator T199 ranks 10th with a log *ZP* score of 0.15.

Table IV. Residue POOL Rankings, POOL log ZP Score, and Change in Catalytic Efficiency for *E. coli* Alkaline Phosphatase Variants

Variant	Shell	Pool rank	POOL log *ZP* score	$(k_{cat}/K_M)_{wild\ type}/(k_{cat}/K_M)_{mutant}$	Reference
D51N	First	1	2.8	12	66
D369N	First	2	2.6	95	19
D327N	First	4	2.5	100	67
D327A	First	4	"	>1,000,000	67
H412N	First	6	2.0	130	68
H331E	First	7	1.2	972	65
D101S	First	8	1.0	0.2	63
H372A	Second	9	1.0	0.27	18
H372D	Second	9	"	3.3	This work
H372L	Second	9	"	2.5	This work
D153G	First	10	0.72	0.2	63
D153N	First	10	"	1.1	64
R166A	First	11	0.60	313	69
K328A	First	13	0.47	3.8	70
H86L	Second	14	0.45	3.2	This work
E57Q	Third	17	0.17	1.8	This work
T155M	First	18	0.16	678	19
S102A	First	19	0.16	150,000	71
M53A	Second	21	0.13	1.5	This work
M53T	Second	21	"	2.4	This work
D330N	Second	22	0.06	0.2	21
D330N	Second	22	"	2.1	This work
Q435E	Second	44	−0.03	1.7	This work
T100V	Second	51	−0.03	0.3	20
V99A	Second	100	−0.06	0.2	20
E150Q	Second	106	−0.06	3.7	This work
S105A	Second	136	−0.06	2.1	This work

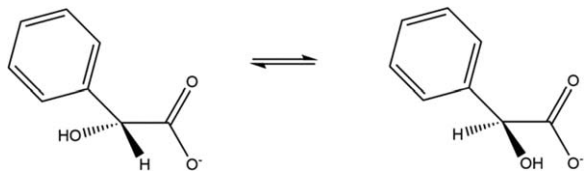
Residue numbering follows the PDB file 1ALK.

Discussion

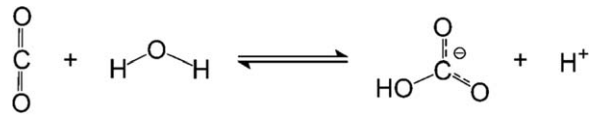
Distal residues can be important contributors to enzyme catalysis but their participation is hard to predict computationally by prior methods. Distal residues are clearly important in computationally driven enzyme design, but better understanding of how to use them is needed.[83] While there are presently few enzymes with extensive data on variants with substitutions in distal positions, the examples presented here suggest that the log *ZP* POOL scores can predict the distal residues and provide useful guidance about the extent of distal residue participation.

The log *ZP* scores provide a more accurate prediction of distal residue participation than the previously utilized 8 to 10% cutoff from the top of the POOL rank order. For most instances, the present log *ZP* score cutoff (log $ZP > 0$) generates fewer false positives than the 8 to 10% cutoff. More importantly, the log *ZP* method allows that some proteins have much more extensive distal residue participation than others.

The multilayer active sites of *P. putida* nitrile hydratase and of human phosphoglucose isomerase are predicted by the log *ZP* scores, as is the single-layer active site of *P. putida* ketosteroid isomerase. The log *ZP* score cutoff utilized here results in overprediction of distal residue involvement in *E. coli* alkaline phosphatase. This may be because of extensive collective interactions in AP, such that mutation at a single position may have no significant effect on catalysis because a sufficient number of additional important residues remain. We note that some of the positions predicted by POOL, such as D153 and D330, for which single-point variants show little difference from wild type, have been shown to be significant in two- and three-point variants.[21] While fewer experimental data points are available for *P. putida* mandelate racemase and for human carbonic anhydrase II, the POOL log *ZP* scores properly predict the previously reported participation of distal residues.



Scheme 1. Reaction catalyzed by mandelate racemase.



Scheme 2. Reaction catalyzed by carbonic anhydrase.

For ketosteroid isomerase, the two residues contributing the most to catalysis, D40 and Y16, are both ranked in the top four by POOL. The 66-fold loss in catalytic efficiency observed[55] for the Y57S variant suggests that the aromatic ring of this residue, ranked second by POOL, is important for catalysis. From very recent X-ray solution scattering and NMR studies of the Y57S variant it was concluded that, while the overall size of the variant structure is reduced relative to wild type, the active site cavity is enlarged,[84] suggesting that the bulky phenyl group is necessary to maintain the correct spatial arrangement within the active site. The number 1 POOL ranking for Y32 is an apparent false positive. It is possible that the very intense electric field[85] inside the active site is not significantly diminished by the loss of one residue in the second shell. Another possible explanation is that, upon mutation of Tyr to Phe, a water molecule takes the place of the phenolic hydroxyl group and assumes its role, perhaps then allowing catalytic efficiency to be maintained.

The log *ZP* scores provide richer information about the importance of amino acid residues in the function of the enzyme than do other computational predictors, including the simple POOL ranks used previously. The cutoff utilized in the present work corresponds to the best performance for the admittedly small available set of proteins with experimental mutagenesis data on distal residues. More examples with extensive experimental data on distal residue mutagenesis are needed to establish the optimum cutoff for the log *ZP* scores.

Distal residues influence the polarity, proton transfer properties, electrostatic ambience, spatial arrangement, and flexibility of active site residues. While the present method predicts most of the distal residues that influence catalysis, it is unable to predict the effects of some substitutions that alter the orientation or relative positions of the catalytic residues by virtue of their steric bulk, for instance in the M105A variant of ketosteroid isomerase.[24]

Since distal residues can play key supporting roles in catalysis, and since the electrostatic properties of the residues in and around the active site are critically important, these factors need to be considered in rational protein design.

Materials and Methods

Computational methods

The protein structures used as the input data for all calculations were downloaded from the Protein Data Bank (PDB, http://www.rcsb.org/pdb/).[86] Coordinates for all the proteins in the test set were analyzed by Theoretical Microscopic Titration Curves (THEMATICS)[5,87] using the method of Wei.[7] POOL calculations[3] were performed as described previously.[4] Unless otherwise noted, input features include THEMATICS metrics, INTREPID evolutionary scores[88] and ConCavity[10] surface geometry features. INTREPID scores were obtained using the Berkeley phylogenomics web server.[89] The structure-only version of ConCavity was used.

The Ligand Protein Contacts—Contacts of Structural Units (LPC-CSU) server was used to identify the ligand- and/or metal-binding residues and the residue-residue contacts,[90,91] in order to assign residues to shells. Both the known catalytic and ligand/metal binding residues were considered first-shell for this study.

Experimental mutagenesis data reported previously were obtained from individual literature references, with the help of the BRENDA enzyme database[92] and from the Protein Mutant Database (http://www.genome.ad.jp/dbget-bin/www_bfind?pmd) within the DBGET database retrieval system (http://www.genome.jp/dbget/).

Nitrile hydratase

Expression, purification, and kinetics analysis were performed exactly as described previously.[23]

Alkaline phosphatase

Site-directed mutagenesis. An expression plasmid pEK29[69] generously provided by E.R. Kantrowitz (Boston College), encoding the *E. coli* alkaline phosphatase gene, was used for protein expression and mutagenesis. Mutations were constructed using a Quikchange site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA). The mutated genes were sequenced to verify the construct (Massachusetts General Hospital DNA core, Cambridge, MA). After confirmation of the intended mutation, the plasmid was transformed into SM547 (a gift from E.R. Kantrowitz)[69] competent cells and selected on LB agar containing 100 μg/mL ampicillin.

Protein expression and purification. Cells were grown at 37°C in 1 L of 2× YT media containing ampicillin (100 μg/mL) for 12 h. The cells were harvested, washed, and osmotically shocked as previously described[93] and then protein was precipitated, suspended, dialyzed, and purified on a HiTrap Fast-Flow Q column (GE Healthcare) as previously described.[69] Protein fractions greater than 95% pure as judged by SDS-PAGE and Coomassie blue staining were pooled and stored at −20°C. The concentration of all proteins was determined by Bradford assay (Bio-Rad, Hercules, CA) using bovine serum albumin as a standard.

Kinetics. Alkaline phosphatase activity was measured spectrophotometrically utilizing *p*-nitrophenyl phosphate as the substrate. Formation of *p*-

nitrophenolate was monitored at 410 nm at room temperature in Tris buffer (1.0*M* Tris-HCl pH 8.0). Initial velocities were calculated utilizing a molar absorptivity coefficient of 1.42 x $10^4 M^{-1}$ cm^{-1}.[69] Nonlinear regression was performed to calculate K_M and k_{cat} values using GraphPad Prism 5 version 5.02.

Acknowledgments

The authors acknowledge the students in the Principles of Chemical Biology laboratory section at Northeastern University and Michelle Silva, Teaching Assistant, for construction of some of the AP variants.

Disclosure: MJO has a grant from MathWorks, Inc. in support of another project.

References

1. Porter CT, Bartlett GJ, Thornton JM (2004) The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 32:D129–D133.
2. Lee J, Goodey NM (2011) Catalytic contributions from remote regions of enzyme structure. Chem Rev 111: 7595–7624.
3. Tong W, Wei Y, Murga LF, Ondrechen MJ, Williams RJ (2009) Partial order optimum likelihood (POOL): maximum likelihood prediction of protein active site residues using 3D structure and sequence properties. PLoS Comp Biol 5:e1000266.
4. Somarowthu S, Yang H, Hildebrand DGC, Ondrechen MJ (2011) High-performance prediction of functional residues in proteins with machine learning and computed input features. Biopolymers 95:390–400.
5. Ondrechen MJ, Clifton JG, Ringe D (2001) THEMATICS: a simple computational predictor of enzyme function from structure. Proc Natl Acad Sci USA 98: 12473–12478.
6. Ko J, Murga LF, André P, Yang H, Ondrechen MJ, Williams RJ, Agunwamba A, Budil DE (2005) Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. Proteins 59:183–195.
7. Wei Y, Ko J, Murga L, Ondrechen MJ (2007) Selective prediction of interaction sites in protein structures with THEMATICS. BMC Bioinform 8:119.
8. Somarowthu S, Ondrechen MJ (2012) POOL server: machine learning application for functional site prediction in proteins. Bioinformatics 28:2078–2079.
9. Hildebrand DGC, Yang H, Ondrechen MJ, Williams RJ (2010) High conservation of amino acids with anomalous protonation behavior. Curr Bioinform 5:134–140.
10. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol 5:e1000585.
11. Sankararaman S, Sjolander K (2008) INTREPID: INformation-theoretic TREe traversal for Protein functional site IDentification. Bioinformatics 24:2445–2452.
12. Johannes TW, Zhao H (2006) Directed evolution of enzymes and biosynthetic pathways. Curr Opin Microbiol 9:261–267.
13. Gould SM, Tawfik DS (2005) Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. Biochemistry 44:5444–5452.
14. Hsu C-C, Hong Z, Wada M, Franke D, Wong C-H (2005) Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. Proc Natl Acad Sci USA 102:9122–9126.
15. Oue S, Okamoto A, Yano T, 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.
16. van den Heuvel RH, van den Berg WA, Rovida S, van Berkel WJ (2004) Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. J Biol Chem 279: 33492–33500.
17. Tomatis PE, Rasia RM, Segovia L, Vila AJ (2005) Mimicking natural evolution in metallo-beta-lactamases through second-shell ligand mutations. Proc Natl Acad Sci USA 102:13761–13766.
18. Xu X, Qin XQ, Kantrowitz ER (1994) Probing the role of histidine-372 in zinc binding and the catalytic mechanism of *Escherichia coli* alkaline phosphatase by site-specific mutagenesis. Biochemistry 33:2279–2284.
19. Hehir MJ, Murphy JE, Kantrowitz ER (2000) Characterization of heterodimeric alkaline phosphatases from *Escherichia coli*: an investigation of intragenic complementation. J Mol Biol 304:645–656.
20. Mandecki W, Shallcross MA, Sowadski J, Tomazic-Allen S (1991) Mutagenesis of conserved residues within the active site of *Escherichia coli* alkaline phosphatase yields enzymes with increased kcat. Protein Eng 4:801–804.
21. Muller BH, Lamoure C, Le Du MH, Cattolico L, Lajeunesse E, Lemaitre F, Pearson A, Ducancel F, Menez A, Boulain JC (2001) Improving *Escherichia coli* alkaline phosphatase efficacy by additional mutations inside and outside the catalytic pocket. Chembiochem 2:517–523.
22. Schafer SL, Barrett WC, Kallarakal AT, Mitra B, Kozarich JW, Gerlt JA, Clifton JG, Petsko GA, Kenyon GL (1996) Mechanism of the reaction catalyzed by mandelate racemase: structure and mechanistic properties of the D270N mutant. Biochemistry 35:5662–5669.
23. Brodkin HR, Novak WR, Milne AC, D'Aquino JA, Karabacak NM, Goldberg IG, Agar JN, Payne MS, Petsko GA, Ondrechen MJ, Ringe D (2011) Evidence of the participation of remote residues in the catalytic activity of Co-type nitrile hydratase from *Pseudomonas putida*. Biochemistry 50:4923–4935.
24. Somarowthu S, Brodkin HR, D'Aquino JA, Ringe D, Ondrechen MJ, Beuning PJ (2011) A tale of two isomerases: compact versus extended active sites in ketosteroid isomerase and phosphoglucose isomerase. Biochemistry 50:9283–9295.
25. Walsh JM, Parasuram R, Rajput PR, Rozners E, Ondrechen MJ, Beuning PJ (2012) Effects of non-catalytic, distal amino acid residues on activity of *E. coli* DinB (DNA polymerase IV). Environ Mol Mutagen 53:766–776.
26. Kobayashi M, Shimizu S (1998) Metalloenzyme nitrile hydratase: structure, regulation, and application to biotechnology. Nat Biotechnol 16:733–736.
27. Hashimoto Y, Sasaki S, Herai S, Oinuma K, Shimizu S, Kobayashi M (2002) Site-directed mutagenesis for cysteine residues of cobalt-containing nitrile hydratase. J Inorg Biochem 91:70–77.
28. Miyanaga A, Fushinobu, S., Ito, K., Shoun, H. Wakagi, T. (2004) Mutational and structural analysis of cobalt-containing nitrile hydratase on substrate and metal binding. Eur J Biochem 271:429–438.

29. Takarada H, Kawano Y, Hashimoto K, Nakayama H, Ueda S, Yohda M, Kamiya N, Dohmae N, Maeda M, Odaka M (2006) Mutational study on alphaGln90 of Fe-type nitrile hydratase from *Rhodococcus sp*. N771. Biosci Biotechnol Biochem 70:881–889.
30. Endo I, Nojiri M, Tsujimura M, Nakasako M, Nagashima S, Yohda M, Odaka M (2001) Fe-type nitrile hydratase. J Inorg Biochem 83:247–253.
31. Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci 24:8–11.
32. Chaput M, Claes V, Portetelle D, Cludts I, Cravador A, Burny A, Gras H, Tartar A (1988) The neurotrophic factor neuroleukin is 90% homologous with phosphohexose isomerase. Nature 332:454–455.
33. Faik P, Walker JIH, Redmill AAM, Morgan MJ (1988) Mouse glucose-6-phosphate isomerase and neuroleukin have identical 3′ sequences. Nature 332:455–456.
34. Watanabe H, Takehana K, Date M, Shinozaki T, Raz A (1996) Tumor cell autocrine motility factor Is the neuroleukin/phosphohexose isomerase polypeptide. Cancer Res 56:2960–2963.
35. Xu W, Seiter K, Feldman E, Ahmed T, Chiao J (1996) The differentiation and maturation mediator for human myeloid leukemia cells shares homology with neuroleukin or phosphoglucose isomerase. Blood 87: 4502–4506.
36. Beutler E, West C, Britton HA, Harris J, Forman L (1997) Glucosephosphate isomerase (GPI) deficiency mutations associated with hereditary nonspherocytic hemolytic anemia (HNSHA). Blood Cells Mol Dis 23: 402–409.
37. Yanagawa T, Funasaka T, Tsutsumi S, Watanabe H, Raz A (2004) Novel roles of the autocrine motility factor/phosphoglucose isomerase in tumor malignancy. Endocr Relat Cancer 11:749–759.
38. Read J, Pearce J, Li X, Muirhead H, Chirgwin J, Davies C (2001) The crystal structure of human phosphoglucose isomerase at 1.6 Å resolution: implications for catalytic mechanism, cytokine activity and haemolytic anaemia. J Mol Biol 309:447–463.
39. Meng M, Chen YT, Hsiao YY, Itoh Y, Bagdasarian M (1998) Mutational analysis of the conserved cationic residues of *Bacillus stearothermophilus* 6-phosphoglucose isomerase. Eur J Biochem 257:500–505.
40. Meng M, Lin HY, Hsieh CJ, Chen YT (2001) Functions of the conserved anionic amino acids and those interacting with the substrate phosphate group of phosphoglucose isomerase. FEBS Lett 499:11–14.
41. Jeffery CJ, Bahnson BJ, Chien W, Ringe D, Petsko GA (2000) Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator. Biochemistry 39:955–964.
42. Solomons JTG, Zimmerly EM, Burns S, Krishnamurthy N, Swan MK, Krings S, Muirhead H, Chirgwin J, Davies C (2004) The crystal structure of mouse phosphoglucose isomerase at 1.6 angstrom resolution and its complex with glucose 6-phosphate reveals the catalytic mechanism of sugar ring opening. J Mol Biol 342:847–860.
43. Lee JH, Jeffery CJ (2005) The crystal structure of rabbit phosphoglucose isomerase complexed with D-sorbitol-6-phosphate, an analog of the open chain form of D-glucose-6-phosphate. Protein Sci 14:727–734.
44. Kraut DA, Sigala PA, Fenn TD, Herschlag D (2010) Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole. Proc Natl Acad Sci 107:1960–1965.
45. Sigala PA, Caaveiro JMM, Ringe D, Petsko GA, Herschlag D (2009) Hydrogen bond coupling in the ketosteroid isomerase active site. Biochemistry 48: 6932–6939.
46. Kraut DA, Churchill MJ, Dawson PE, Herschlag D (2009) Evaluating the potential for halogen bonding in the oxyanion hole of ketosteroid isomerase using unnatural amino acid mutagenesis. ACS Chem Biol 4: 269–273.
47. Chakravorty DK, Soudackov AV, 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.
48. Warshel A, Sharma PK, Chu ZT, Åqvist 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.
49. Kraut DA, Sigala PA, Pybus B, Liu CW, Ringe D, Petsko GA, Herschlag D (2006) Testing electrostatic complementarity in enzyme catalysis: hydrogen bonding in the ketosteroid isomerase oxyanion hole. PLoS Biol 4:e99.
50. Hanoian P, Sigala PA, Herschlag D, Hammes-Schiffer S (2010) Hydrogen bonding in the active site of ketosteroid isomerase: electronic inductive effects and hydrogen bond coupling. Biochemistry 49:10339–10348.
51. Schwans JP, Kraut DA, Herschlag D (2009) Determining the catalytic role of remote substrate binding interactions in ketosteroid isomerase. Proc Natl Acad Sci USA 106:14271–14275.
52. Kim SW, Cha SS, Cho HS, Kim JS, Ha NC, Cho MJ, Joo S, Kim KK, Choi KY, Oh BH (1997) High-resolution crystal structures of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue. Biochemistry 36:14030–14036.
53. Cho HS, Choi G, Choi KY, Oh BH (1998) Crystal structure and enzyme mechanism of Delta 5-3-ketosteroid isomerase from *Pseudomonas testosteroni*. Biochemistry 37:8325–8330.
54. Kim D-H, Jang DS, Nam GH, Choi G, Kim J-S, Ha N-C, Kim M-S, Oh B-H, Choi KY (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.
55. Nam GH, Jang DS, Cha SS, Lee TH, Kim DH, Hong BH, Yun YS, Oh BH, Choi KY (2001) Maintenance of alpha-helical structures by phenyl rings in the active-site tyrosine triad contributes to catalysis and stability of ketosteroid isomerase from *Pseudomonas putida* biotype B. Biochemistry 40:13529–13537.
56. Kim S, Choi K (1995) Identification of active site residues by site-directed mutagenesis of delta 5-3-ketosteroid isomerase from *Pseudomonas putida* biotype B. J Bacteriol 177:2602–2605.
57. Kim DH, Nam GH, Jang DS, Choi G, Joo S, Kim JS, Oh BH, Choi KY (1999) Roles of active site aromatic residues in catalysis by ketosteroid isomerase from *Pseudomonas putida* biotype B. Biochemistry 38: 13810–13819.
58. Kim S, Joo S, Choi G, Cho H, Oh B, Choi K (1997) Mutational analysis of the three cysteines and active-site aspartic acid 103 of ketosteroid isomerase from *Pseudomonas putida* biotype B. J Bacteriol 179:7742–7747.

59. Kim EE, Wyckoff HW (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. J Mol Biol 218:449–464.
60. Andrews LD, Zalatan JG, Herschlag D (2014) Probing the origins of catalytic discrimination between phosphate and sulfate monoester hydrolysis: comparative analysis of alkaline phosphatase and protein tyrosine phosphatases. Biochemistry 53:6811–6819.
61. O'Brien PJ, Lassila JK, Fenn TD, Zalatan JG, Herschlag D (2008) Arginine coordination in enzymatic phosphoryl transfer: evaluation of the effect of Arg166 mutations in *Escherichia coli* alkaline phosphatase. Biochemistry 47:7663–7672.
62. Coleman JE (1992) Structure and mechanism of alkaline phosphatase. Annu Rev Biophys Biomol Struct 21: 441–483.
63. Dealwis CG, Chen L, Brennan C, Mandecki W, Abad-Zapatero C (1995) 3-D structure of the D153G mutant of *Escherichia coli* alkaline phosphatase: an enzyme with weaker magnesium binding and increased catalytic activity. Protein Eng 8:865–871.
64. Matlin AR, Kendall DA, Carano KS, Banzon JA, Klecka SB, Solomon NM (1992) Enhanced catalysis by active-site mutagenesis at aspartic acid 153 in *Escherichia coli* alkaline phosphatase. Biochemistry 31: 8196–8200.
65. Wojciechowski CL, Kantrowitz ER (2003) Glutamic acid residues as metal ligands in the active site of *Escherichia coli* alkaline phosphatase. Biochim Biophys Acta 1649:68–73.
66. Tibbitts TT, Murphy JE, Kantrowitz ER (1996) Kinetic and structural consequences of replacing the aspartate bridge by asparagine in the catalytic metal triad of *Escherichia coli* alkaline phosphatase. J Mol Biol 257: 700–715.
67. Tibbitts TT, Xu X, Kantrowitz ER (1994) Kinetics and crystal structure of a mutant *Escherichia coli* alkaline phosphatase (Asp-369 → Asn): a mechanism involving one zinc per active site. Protein Sci 3:2005–2014.
68. Ma L, Tibbitts TT, Kantrowitz ER (1995) *Escherichia coli* alkaline phosphatase: X-ray structural studies of a mutant enzyme (His-412–>Asn) at one of the catalytically important zinc binding sites. Protein Sci 4:1498–1506.
69. Chaidaroglou A, Brezinski DJ, Middleton SA, Kantrowitz ER (1988) Function of arginine-166 in the active site of *Escherichia coli* alkaline phosphatase. Biochemistry 27:8338–8343.
70. Xu X, Kantrowitz ER (1991) A water-mediated salt link in the catalytic site of *Escherichia coli* alkaline phosphatase may influence activity. Biochemistry 30:7789–7796.
71. Stec B, Hehir MJ, Brennan C, Nolte M, Kantrowitz ER (1998) Kinetic and X-ray structural studies of three mutant *E. coli* alkaline phosphatases: insights into the catalytic mechanism without the nucleophile Ser102. J Mol Biol 277:647–662.
72. Mitra B, Kallarakal AT, Kozarich JW, Gerlt JA, Clifton JG, Petsko GA, Kenyon GL (1995) Mechanism of the reaction catalyzed by mandelate racemase: importance of electrophilic catalysis by glutamic acid 317. Biochemistry 34:2777–2787.
73. Landro JA, Kallarakal AT, Ransom SC, Gerlt JA, Kozarich JW, Neidhart DJ, Kenyon GL (1991) Mechanism of the reaction catalyzed by mandelate racemase. 3. Asymmetry in reactions catalyzed by the H297N mutant. Biochemistry 30:9274–9281.
74. Kallarakal AT, Mitra B, Kozarich JW, Gerlt JA, Clifton JG, Petsko GA, Kenyon GL (1995) Mechanism of the reaction catalyzed by mandelate racemase: structure and mechanistic properties of the K166R mutant. Biochemistry 34:2788–2797.
75. Neidhart DJ, Howell PL, Petsko GA, Powers VM, Li RS, Kenyon GL, Gerlt JA (1991) Mechanism of the reaction catalyzed by mandelate racemase. 2. Crystal structure of mandelate racemase at 2.5-A resolution: identification of the active site and possible catalytic residues. Biochemistry 30:9264–9273.
76. Lesburg CA, Huang C, Christianson DW, Fierke CA (1997) Histidine –> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase II alter metal coordination geometry but retain catalytic activity. Biochemistry 36:15780–15791.
77. Christianson DW, Cox JD (1999) Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. Annu Rev Biochem 68:33–57.
78. Kiefer LL, Krebs JF, Paterno SA, Fierke CA (1993) Engineering a cysteine ligand into the zinc binding site of human carbonic anhydrase II. Biochemistry 32: 9896–9900.
79. Eriksson AE, Jones TA, Liljas A (1988) Refined structure of human carbonic anhydrase II at 2.0 A resolution. Proteins 4:274–282.
80. Ippolito JA, Christianson DW (1993) Structure of an engineered His3Cys zinc binding site in human carbonic anhydrase II. Biochemistry 32:9901–9905.
81. Venta PJ, Welty RJ, Johnson TM, Sly WS, Tashian RE (1991) Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His——Tyr): complete structure of the normal human CA II gene. Am J Hum Genet 49:1082–1090.
82. Liang Z, Xue Y, Behravan G, Jonsson BH, Lindskog S (1993) Importance of the conserved active-site residues Tyr7, Glu106 and Thr199 for the catalytic function of human carbonic anhydrase II. Eur J Biochem 211:821–827.
83. Hilvert D (2013) Design of protein catalysts. Annu Rev Biochem 82:447–470.
84. Cha HJ, Jang DS, Jin KS, Lee HJ, B.H. H, Kim E-S, Kim J, Lee HC, Choi KY, Ree M (2014) Three-dimensional structures of a wild-type ketosteroid isomerase and its single mutant in solution. Sci Advan Mater 6: 2325–2333.
85. Fried SD, Bagchi S, Boxer SG (2014) Extreme electric fields power catalysis in the active site of ketosteroid isomerase. Science 346:1510–1514.
86. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.
87. Ondrechen MJ. Identification of functional sites based on prediction of charged group behavior. In: Baxevanis AD, Davison DB, Page RDM, Petsko GA, Stein LD, Stormo GD, Eds. (2004) Current protocols in bioinformatics. Hoboken, NJ: Wiley, pp 8.6.1–8.6.10.
88. Sankararaman S, Kolaczkowski B, Sjolander K (2009) INTREPID: a web server for prediction of functionally important residues by evolutionary analysis. Nucleic Acids Res 37:W390–W395.
89. Glanville JG, Kirshner D, Krishnamurthy N, Sjolander K (2007) Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis. Nucleic Acids Res 35:W27–W32.
90. Sobolev V, Eyal E, Gerzon S, Potapov V, Babor M, Prilusky J, Edelman M (2005) SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. Nucleic Acids Res 33:W39–W43.

91. Sobolev V, Sorokine A, Prilusky J, Abola E, Edelman M (1999) Automated analysis of interatomic contacts in proteins. Bioinformatics 15:327–332.
92. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res 37:D588–592.
93. Brockman RW, Heppel LA (1968) On the localization of alkaline phosphatase and cyclic phosphodiesterase in *Escherichia coli*. Biochemistry 7:2554–2562.