

FULL-LENGTH PAPER



Electrostatic fingerprints of catalytically active amino acids in enzymes

Suhasini M. Iyengar | Kelly K. Barnsley | Rhoe Xu | Aleksandr Prystupa | Mary Jo Ondrechen

Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts, USA

Correspondence

Mary Jo Ondrechen, Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA.
Email: mjo@neu.edu

Present address

Rhoe Xu, Bioinformatics and Computational Biology Program, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

Funding information

Fulbright Research Grant, Grant/Award Number: 2221103; National Science Foundation, Grant/Award Number: CHE-1905214

Review Editor: Nir Ben-Tal

Abstract

The computed electrostatic and proton transfer properties are studied for 20 enzymes that represent all six major enzyme commission classes and a variety of different folds. The properties of aspartate, glutamate, and lysine residues that have been previously experimentally determined to be catalytically active are reported. The catalytic aspartate and glutamate residues studied here are strongly coupled to at least one other aspartate or glutamate residue and often to multiple other carboxylate residues with intrinsic pK_a differences less than 1 pH unit. Sometimes these catalytic acidic residues are also coupled to a histidine residue, such that the intrinsic pK_a of the acidic residue is higher than that of the histidine. All catalytic lysine residues studied here are strongly coupled to tyrosine or cysteine residues, wherein the intrinsic pK_a of the anion-forming residue is higher than that of the lysine. Some catalytic lysines are also coupled to other lysines with intrinsic pK_a differences within 1 pH unit. Some evidence of the possible types of interactions that facilitate nucleophilicity is discussed. The interactions reported here provide important clues about how side chain functional groups that are weak Brønsted acids or bases for the free amino acid in solution can achieve catalytic potency and become strong acids, bases or nucleophiles in the enzymatic environment.

KEYWORDS

amino acid interactions, enzyme catalysis, enzyme mechanisms, protein electrostatics

1 | INTRODUCTION

Enzymes catalyze reactions under mild conditions that on the benchtop might require extreme conditions such as high temperature or strong acid or base; how this is achieved is a topic of intense current interest.^{1,2} Enzyme

catalysis often requires exceptional chemical properties for amino acid side chains that, in the absence of the protein environment, would be far less reactive. For instance, amino acid side chains that would be weak Brønsted acids or bases in a small peptide can, in the right protein environment, become a strong base that

Abbreviations: ACCD, 1-aminocyclopropane-1-carboxylate deaminase; DHDPs, dihydrodipicolinate synthase; DNase, deoxyribonucleotidase; EC, enzyme commission; FAD, flavin adenine dinucleotide; FAH, fumarylacetoacetate hydrolase; HAD, haloacid dehalogenase; HGPRT, hypoxanthine-guanine phosphoribosyltransferase; M-CSA, mechanism and Catalytic Site Atlas; OCD, ornithine cyclodeaminase; PGI, phosphoglucose isomerase; SepSecS, O-phosphoserine-tRNA:selenocysteine-tRNA synthase; TH, pyrogallol hydroxytransferase, also called transhydroxylase.

deprotonates a C—H bond or a strong acid that protonates the hydroxyl group of an alcohol. The primary amine side chain of a lysine residue, that would normally be protonated at neutral pH, can be deprotonated by the enzymatic environment to serve as a nucleophile. In this work, the types of interactions that impart these extraordinary chemical properties are explored, to provide insight into how amino acid side chains can become highly reactive in the enzymatic ambience. Here, 20 enzymes with literature annotations of residue roles in catalysis are analyzed using properties calculated from the proton occupation functions, which are in turn obtained from the computed electrostatic potential. These enzymes were chosen to include all six major enzyme commission (EC) classes and a variety of different folds (Table S1).

One chemical property that facilitates catalysis in enzymes is an expanded buffer range of the ionizable residues involved in catalysis. Indeed, this is such a universal property of biochemically active residues that we have used it successfully to identify the reactive amino acids in protein structures.^{3–8} This expanded buffer range is a simple polyprotic acid effect arising from interactions between residues that can transfer protons. The expanded buffer range enables both protonation states to exist over a wide pH range, so that the active amino acids can return to their original state for the next turnover cycle. Reactivity is only probable within the buffer range of the titration curve and, in a dynamic environment, an expanded buffer range is advantageous, and apparently essential, for catalysis.

The computed proton occupation function $C(\text{pH})$ describes the theoretical protonation equilibrium of an amino acid in a protein structure across a wide pH range. $C(\text{pH})$ may be calculated for each residue with proton transfer capability, using the computed electrical potential function^{9,10} for the protein structure. For a residue that follows Henderson–Hasselbalch behavior, the function $C(\text{pH})$, which always equals 1 at low pH and zero at high pH, is a sigmoidal function with a sharp fall-off at pH values near the $\text{p}K_a$. For biochemically active amino acids, such as catalytic residues or those involved in ligand binding, C has anomalous shape,¹¹ resulting from strong coupling to proton transfer equilibria on nearby residues, as is observed in polyprotic acids. It is useful to express the $C(\text{pH})$ functions in terms of a first derivative function f ,^{3,4} defined as:

$$f(\text{pH}) = -dC/d(\text{pH}) \quad (1)$$

The $f(\text{pH})$ function is related to the proton binding capacity,¹² a measure of the change in concentration of a bound species per unit change in its chemical potential.

For a titratable amino acid side chain in a small peptide that does not interact with other titratable species, the $f(\text{pH})$ function resembles a single Gaussian distribution function, with a peak at or near the pH equal to the $\text{p}K_a$. However, for titratable residues involved in catalysis or ligand binding, the broadened $f(\text{pH})$ function deviates substantially from Gaussian form and may be asymmetrical or have multiple peaks.^{3,4} The $f(\text{pH})$ function therefore resembles a probability density function which, when integrated over all real numbers, yields unity because of the 1 to 0 range of the $C(\text{pH})$ functions. The shape of the $f(\text{pH})$ function can be described by its n th central moments μ_n , defined as:⁴

$$\mu_n = \int (\text{pH} - m_1)^n [f(\text{pH})] d(\text{pH}), \quad (2)$$

where m_1 is the first moment, defined for $n = 1$ by the expression for the n th moment m_n as:

$$m_n = \int (\text{pH})^n [f(\text{pH})] d(\text{pH}), \quad (3)$$

where the integrals in Equations (2) and (3) are over all real numbers $(-\infty \text{ to } +\infty)$.⁴ The first central moment μ_1 is related to the $\text{p}K_a$, the second central moment μ_2 is the variance, μ_3 is the skewness, and μ_4 is the kurtosis of the distribution. The even-numbered central moments are always positive, whereas the odd-numbered central moments can be positive or negative. A large $|\mu_3|$ indicates an asymmetrical $f(\text{pH})$ function and a large μ_4 is indicative of a broadened $f(\text{pH})$ function and a widened buffer range. A large, positive μ_3 indicates broadening on the high pH side of $f(\text{pH})$, with a large, negative μ_3 corresponding to broadening on the low pH side. Large values for $|\mu_3|$ and μ_4 are excellent indicators that a residue is active in catalysis and/or ligand recognition.^{3,4}

2 | RESULTS

Recently,¹³ we reported expressions that define certain types of pairwise interactions between amino acids that help to promote catalysis. The proton transfer equilibria of polyprotic acids¹⁴ depend on the energy of interaction ϵ between the pairs of functional groups and on the $\text{p}K_a$ difference between them. The model $\text{p}K_a$ of an amino acid side chain in a protein structure is defined as the $\text{p}K_a$ of the side chain for the free amino acid in aqueous solution. The intrinsic $\text{p}K_a$ of an amino acid in a protein structure is defined as the $\text{p}K_a$ of the side chain in the hypothetical protein with all other titratable groups in their electrically neutral state. For two like-charge

catalytic glutamates E166 and E351 plus three ordinary glutamates, E227, E234, and E340, of β -glucosidase from *Thermotoga maritima* are shown in Figure 1; the overall reaction is shown in Scheme 1. Note how the curves for the two biochemically active glutamates are shifted to lower pH and how the nucleophile, E351, has an extended buffer range on the high pH side of the midpoint.

The residues most strongly coupled to E166 and E351 for β -glucosidase from *Thermotoga maritima* are listed in Table 2, which lists the pairwise potential energy (in kcal/mol) between each biochemically active glutamate and each coupling partner and the intrinsic pK_a s of

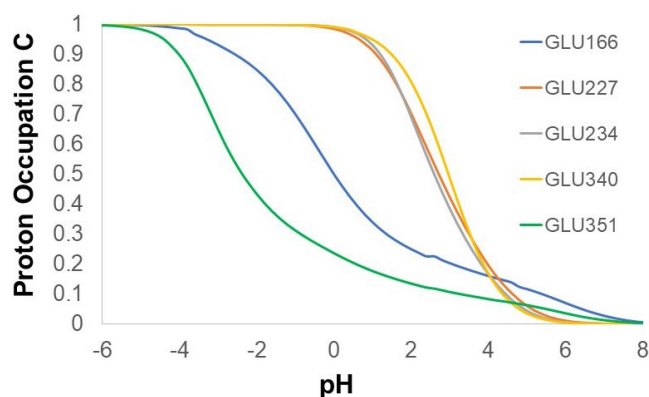


FIGURE 1 Theoretical titration curves, expressed as the proton occupation C for each residue in a large ensemble of protein molecules as a function of pH, calculated for the two catalytic glutamates E166 (acid/base) and E351 (nucleophile) plus three ordinary glutamates, E227, E234, and E340, of β -glucosidase from *Thermotoga maritima*



SCHEME 1 Reaction catalyzed by β -glucosidase, where $R = \beta$ -D-glucose

the coupling partners. The coupling partners, R77, H121, Y295, and E405, all surround the active site pocket (Figure 2). E166 has an intrinsic pK_a of 3.7 and E351 has an intrinsic pK_a of 4.3. E166 is strongly coupled to two other acidic residues, E351 and E405; both have intrinsic pK_a s within 1 pH unit of that of E166 and contribute to the broad buffer range of E166. The protein environment has downshifted the intrinsic pK_a of H121, so that the intrinsic pK_a of E166 is higher and therefore the H121-E166 coupling contributes to a broader titration curve for both residues. While E166 is strongly coupled to its partners, note the even higher energies of interaction between the nucleophilic E351 and its partners. E351 is coupled to E166 and to E405, with differences in intrinsic pK_a of less than 1 pH unit. The potential energy between E351 and H121 is 1.8 kcal/mol = 1.3 in units of $-\ln(10)$ RT (for $T = 293$ K), so the intrinsic pK_a difference between E351 and H121 of 0.9 pH is within the range that leads to broadening of the buffer range. Interactions of both catalytic residues with R77 and Y295 shift the titration curve along the pH axis but do not alter the shape of the titration curve or contribute to the broad

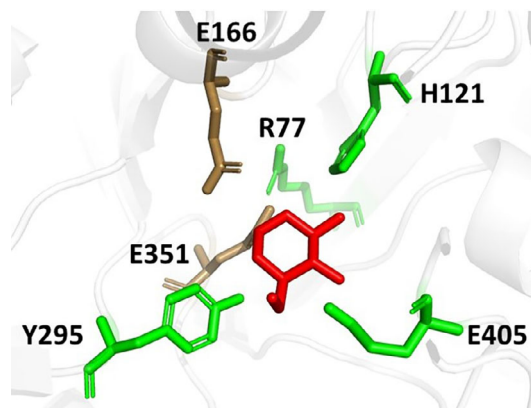


FIGURE 2 Coupled amino acids around the active site of β -glucosidase from *Thermotoga maritima*: Catalytic E166 and E351 (brown); other coupled residues (green). Ligand is shown in red. Based on PDB structure 1OIF.²⁰ Image rendered in Pymol 2.4.1

TABLE 2 Pairwise energies of interaction (kcal/mol) between the two catalytic glutamate residues, E166 (general acid/bas) and E351 (nucleophile) of β -glucosidase from *Thermotoga maritima* and their strongest coupling partners; intrinsic pK_a s are also listed

Top couplers to E166: $pK_a(\text{intrinsic}) = 3.7$			Top couplers to E351: $pK_a(\text{intrinsic}) = 4.3$		
Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$	Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$
E351	1.9	4.3	Y295	3.1	10.6
Y295	1.4	10.6	R77	2.6	10.5
H121	1.3	3.4	E166	1.9	3.7
R77	1.2	10.5	H121	1.8	3.4
E405	0.95	4.3	E405	1.0	4.3

buffer ranges. The residues corresponding to R77 and Y295 in an orthologous enzyme from *Spodoptera frugiperda* (an agricultural pest insect) have been shown previously by site-directed mutagenesis to lower the optimum pH of the reaction.²⁴

Pairwise interaction energies for the two catalytic glutamate residues of the other four glycoside hydrolases listed in Table 1 are given in the Tables S2–S5. For all these examples, the catalytic glutamates are coupled to other acidic residues (aspartates and glutamates) with intrinsic pK_a differences of 1 pH unit or less, to yield broad buffer ranges for the catalytic residues. Note that positively charged residues, including R92 of Exo- β -(1,3)-glucanase, R136 and K179 of Endo-1,4- β -glucanase, R45 of β -1,4-galactanase, and R62 of Endocellulase E1, are coupled to the catalytic residues to affect a lower pK_a and stronger acidity.

A general conclusion from these five glycoside hydrolases is that the nucleophile, compared to other glutamate residues, has a downshifted titration curve (corresponding to the low μ_1 value) with an expanded buffer range, where the titration curve is broadened on the high pH side of the pK_a (corresponding to the high, positive μ_3 value). This is consistent with available experimental evidence. Søndergaard and co-workers reported fitted titration curves of individual residues obtained from experimental C^{13} NMR data for *Bacillus circulans* xylanase (Uniprot P09850; EC 3.2.1.8; PDB 1XNB).²⁵ This 20.4 kDa protein has only two glutamates: E78, the nucleophile and E172, the proton donor.²⁶ Table 3 summarizes the titration curve features for these two residues, as calculated by us and determined by Søndergaard. The pK_a , the buffer range (equal to the number of pH units between 0.10 and 0.90 average proton occupancy) and the number of pH units on the high pH side of the buffer range, between 0.10 and 0.50 average proton occupancy (i.e., between the pH at the high pH end of the buffer range and the pH at the pK_a). These similar data sets show a lower pK_a and a broadened buffer range on the high pH side for the nucleophile. The two catalytic residues E78 and E172 are the principal contributors to the buffer range of one another. The stronger coupling of the nucleophile E78 to the positively charged neighbor R112 creates the lower pK_a for E78 (Table S6);

this ensures that E78 is deprotonated at neutral or near-neutral pH and available to affect nucleophilic attack.

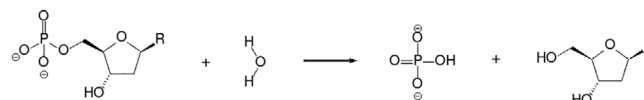
2.1.2 | Human deoxyribonucleotidase

Deoxyribonucleotidase (DNase) degrades DNA by catalyzing the hydrolysis of phosphodiester linkages in the backbone of DNA (Scheme 2). Human mitochondrial deoxyribonucleotidase (Uniprot Q9NPB1; EC. 3.1.3.5; PDB 1Q91) has two catalytic aspartate residues.²⁷ Asp41 serves as the nucleophile, Asp 43 is the proton donor and acceptor; a third active site aspartate, Asp176, coordinates a Mg^{2+} ion.^{28,29} Table 4 gives the pairwise potential energies and the intrinsic pK_a s of the five strongest coupling partners for D41 and D43. It is observed that D41 is coupled with high potential energy to three other acidic residues, D175, D176, and D43, with intrinsic pK_a s very close to that of D41. The side chain of D175, a second shell residue, is located about 3.7 Å away from that of D41 (Figure 3). D43 has strong coupling to two acidic residues, D41 and D176, with very similar intrinsic pK_a s.

Theoretical titration curves for the two catalytic aspartates and for three other aspartates in human deoxyribonucleotidase are shown in Figure 4. In Figure 4, one observes the downshifted titration curves for the two catalytic residues and the long tail on the high pH side of the curve for the nucleophilic D41.

2.1.3 | Haloacid dehalogenase

Haloacid dehalogenase (HAD) catalyzes the hydrolytic dehalogenation of an L-2-haloacid to produce the corresponding D-2-hydroxyacid with inversion at



SCHEME 2 Reaction catalyzed by deoxyribonucleotidase; R = A, G, C, or T

TABLE 3 Comparison of titration curve features of the catalytically active residues of *B. circulans* xylanase, E78 (nucleophile) and E172 (proton donor), as calculated here and as measured from fitted C^{13} NMR chemical shift curves as reported in Ref. 25

	This work—calculated			Ref. 25—from NMR data		
	pK_a	Range (0.1–0.9)	Range (0.1–0.5)	pK_a	Range (0.1–0.9)	Range (0.1–0.5)
E78	4.6	3.2	1.8	4.7	2.7	1.7
E172	5.9	2.8	1.1	6.6	2.5	0.9

Note: Listed are the pK_a , the number of pH units between 0.10 and 0.90 proton occupancy (the buffer range), and the number of pH units between 0.10 and 0.50 proton occupancy (the part of the buffer range on the high pH side of the pK_a).

TABLE 4 Pairwise electrostatic potential energies ($|E|$, kcal/mol) and intrinsic pK_a s for the five residues most strongly coupled to D41 (nucleophile) and D43 (acid/base) in deoxyribonucleotidase

Top couplers to D41: $pK_a(\text{intrinsic}) = 3.9$			Top couplers to D43: $pK_a(\text{intrinsic}) = 3.9$		
Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$	Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$
D175	3.0	4.2	K143	2.8	6.9
K165	2.8	8.7	D41	1.2	3.9
D176	1.7	4.0	C139	1.0	8.6
D43	1.2	3.9	K165	0.97	8.7
K143	1.0	6.9	D176	0.97	4.0

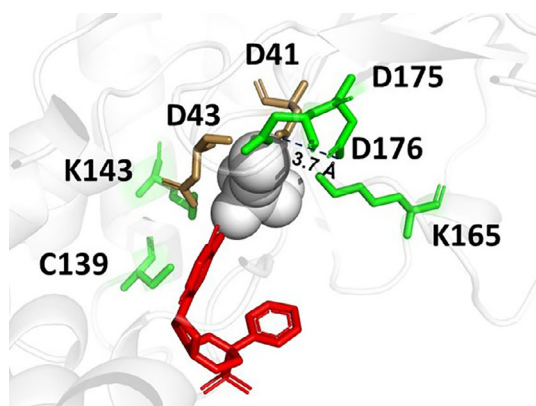


FIGURE 3 Coupled amino acids around the active site of human deoxyribonucleotidase: Catalytic residues D41 and D43 (brown); other coupled residues (green). Ions shown as gray balls. Ligand is shown in red. Based on PDB structure 1Q91.²⁸ Image rendered in Pymol 2.4.1

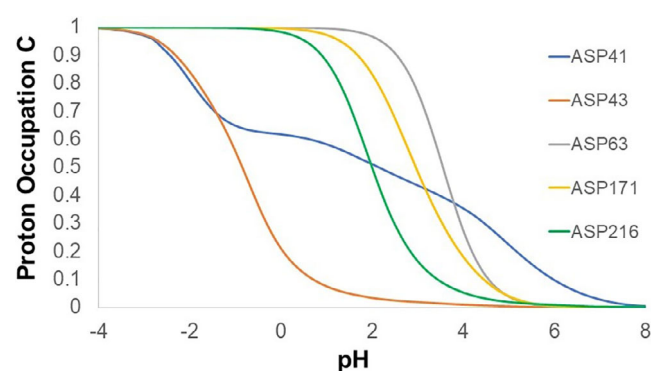
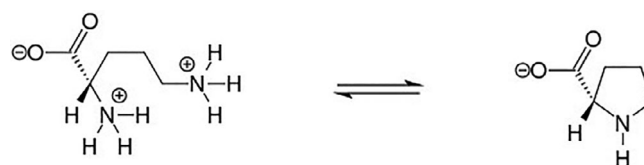


FIGURE 4 Theoretical titration curves, expressed as the proton occupation for each residue in a large ensemble of protein molecules as a function of pH, calculated for D41 (nucleophile), D43 (proton donor/acceptor), and three other more typical aspartates in human DNase

carbon-2.^{30,31} In *Xanthobacter autotrophicus* HAD (Uniprot Q60099; EC 3.8.1.2; PDB 1QQ5), D8 serves as the nucleophile and D176 as the general acid/base. These two residues are strongly coupled to each other with intrinsic pK_a s 0.6 pH units apart, contributing to



SCHEME 3 Reaction catalyzed by ornithine cyclodeaminase

broadened titration curves. Both residues are coupled to the nearby positively charged K147 and R39, leading to downshifted pK_a s, estimated at -2 for D8 and 0.4 for D176. Additional couplings are noted to Y10, Y153, and Y223 (Table S7).

2.1.4 | Ornithine cyclodeaminase

Ornithine cyclodeaminase (OCD) is an NAD^+ -dependent enzyme that catalyzes the conversion of L-ornithine to L-proline and ammonium³² (Scheme 3). The x-ray crystal structure of *Pseudomonas putida* OCD has been reported (Uniprot Q88H32; EC 4.3.1.12; PDB 1X7D).³² D228 is annotated as a general acid/base that deprotonates the α -amino group of ornithine to activate it for nucleophilic attack and then donates the proton back to the α -amino group of the proline product. It is also suggested that Glu56 might function as a proton acceptor. Pairwise electrostatic potential energies and intrinsic pK_a s for the residues most strongly coupled to E56 and D228 are shown in Table 5 and are depicted in Figure 5. E56 and D228 are coupled to each other and to E256, with all pairwise intrinsic pK_a s differences less than 1 pH unit. R45 and K232 are coupled to both E56 and D228 and these interactions with positively charged residues lead to lower pK_a s for the two active site residues.

2.1.5 | Hypoxanthine-guanine phosphoribosyltransferase

Human hypoxanthine-guanine phosphoribosyltransferase (HGPRT) catalyzes the transfer of the phosphoribosyl

TABLE 5 Pairwise electrostatic potential energies ($|E|$, kcal/mol) and intrinsic pK_a s for the five residues most strongly coupled to E56 (possible proton donor) and D228 (general acid/base) in ornithine cyclodeaminase

Top couplers to E56: $pK_a(\text{intrinsic}) = 5.5$			Top couplers to D228: $pK_a(\text{intrinsic}) = 5.2$		
Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$	Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$
R45	2.2	12.0	K232	2.4	9.3
E256	1.3	5.6	R45	1.5	12.0
D228	1.3	5.2	C229	1.4	9.7
C229	0.97	9.7	E56	1.3	5.5
K232	0.92	9.3	E256	0.94	5.6

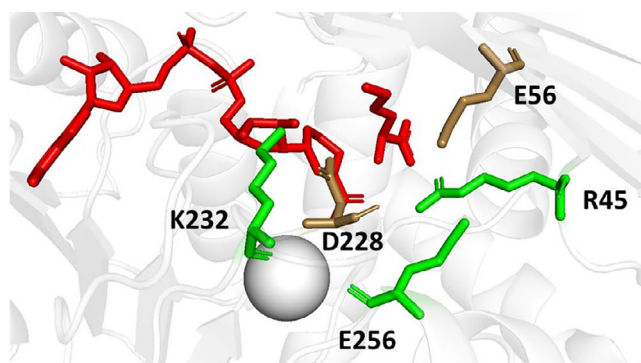


FIGURE 5 Coupled amino acids around the active site of *Pseudomonas putida* ornithine cyclodeaminase: Catalytic residues D228 and E56 (brown); coupled residues (green). Ligands NAD and ornithine are shown in red. Na^+ ion is rendered as a gray ball. Based on PDB structure 1X7D.³² Image rendered in Pymol 2.4.1

group from α -D-5-phosphoribosyl 1-pyrophosphate to hypoxanthine to form the nucleotide IMP or to guanine to form GMP.³³ In human HGPRT (Uniprot P00492; EC 2.4.2.8; PDB 1BZY) D137 serves as the general acid/base, while E133 and D134 help to stabilize the transition state.³⁴ K165 and R169 are strongly coupled to the catalytic D137, serving to downshift its pK_a . D137 has an intrinsic pK_a of 4.7; its buffer range is expanded through coupling to E133 (intrinsic pK_a 4.1) and to D184 (intrinsic pK_a 4.0), a second-shell residue located behind D137 with respect to the substrate (Table S8, Figure 6).

2.1.6 | Aspartate ammonia ligase

Aspartate ammonia ligase catalyzes the ATP-dependent formation of L-asparagine from L-aspartate and ammonia.³⁵ In *E. coli* aspartate ammonia ligase (Uniprot P00963; EC 6.3.1.1; PDB 12AS),³⁶ D46 deprotonates the ammonium ion to activate it for nucleophilic attack. D46 is coupled to D118 with an intrinsic pK_a difference of 0.7 pH units and this interaction serves to expand the buffer

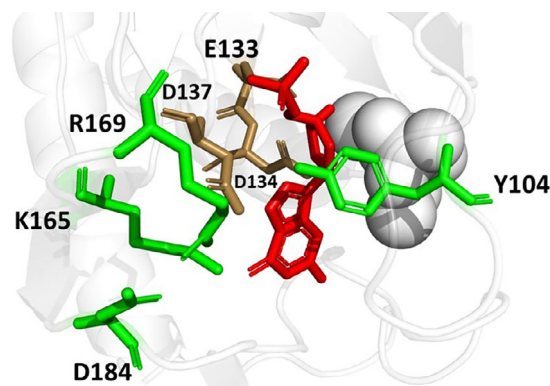


FIGURE 6 Coupled amino acids around the active site of human hypoxanthine-guanine phosphoribosyltransferase (HGPRT). Catalytic residues D137 (general acid/base) and E133 and D134 (transition state stabilizers) are shown in brown; coupled residues are shown in green. A transition state analog inhibitor is shown in red. Pyrophosphate and Mg^{2+} ions are shown as gray balls. Based on PDB structure 1BZY.³³ Image rendered in Pymol 2.4.1

ranges of both residues, which are 6.7 Å apart. The pK_a of the catalytic base D46 is lowered through interactions with the positively charged residues K75 (a second-shell residue) and K77 (a substrate-binding residue) (Table S9).

2.1.7 | Fucose isomerase

E. coli L-fucose isomerase (Uniprot P69922; E.C. 5.3.1.25; PDB 1FUI) converts the aldohexoses L-fucose and L-arabinose to the corresponding ketoses, L-fuculose and L-ribulose, respectively.³⁷ In this manganese-dependent enzyme, E337 and D361 are reported to play general acid/base roles and are involved in binding the Mn^{2+} ion, while H528 also binds the Mn^{2+} ion.³⁷ The intrinsic pK_a s of E337 and D361 are 4.6 and 5.0, respectively, so they are coupled to each other to produce broadened buffer ranges. E337 is also coupled to E316, D313, and D339, with intrinsic pK_a s of 3.7, 3.8, and 4.1, respectively, with the three

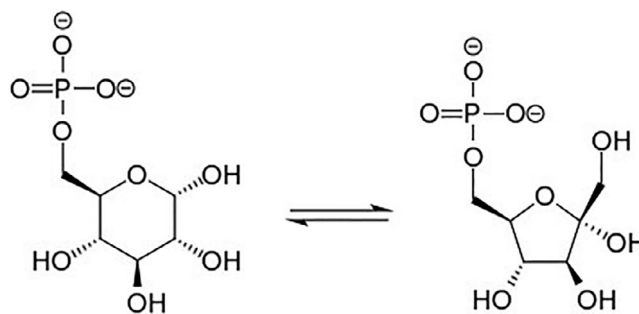
partners contributing to the broadening of the titration curve of E337. Both catalytic residues are coupled to H528; the protein environment has downshifted the intrinsic pK_a of H528 to 3.9, below that of the two catalytic acidic residues, so that H528 contributes to the broadened buffer ranges of E337 and D361 (Table S10).

2.1.8 | Pyrogallol hydroxytransferase

Pyrogallol hydroxytransferase, also called trans-hydroxylase (TH), from the anaerobic bacterium *Pelobacter acidigallici* (Uniprot P80563; EC 1.97.1.2; PDB 4V4E) is an Mo-dependent enzyme that catalyzes a net hydroxyl group transfer, converting pyrogallol to phloroglucinol.³⁸ H144, D174, and Y404 are reported to play general acid/base roles in the mechanism.^{38,39} The active site of TH has an especially extensive network of strongly coupled residues; D174 and H144 each have 12 coupling partners with energies of interaction greater than or equal to 1.0 kcal/mol (Table S11). The buffer range of D174, which has an intrinsic pK_a of 3.2, is broadened through coupling to E177 and E818, with intrinsic pK_a s of 3.5 and 3.7, respectively. The intrinsic pK_a s of the histidine coupling partners of D174, H744 (3.9), H144 (4.3), and H171 (4.1), although decreased by the protein environment, are too high to affect significant broadening of the titration curve of D174. The buffer range of H144, with its intrinsic pK_a of 4.3, is broadened by coupling to other histidine residues with intrinsic pK_a s within 1 pH unit, H744 (3.9) and H738 (3.8).

2.1.9 | Human phosphoglucose isomerase

Phosphoglucose isomerase (PGI) catalyzes the reversible isomerization of glucose 6-phosphate and fructose 6-phosphate (Scheme 4). Human PGI (Uniprot P06744; EC 5.3.1.9; PDB 1IAT) is active as a dimer; E358 acts as the catalytic base, R273 helps to stabilize the intermediate, and H389 and K519 are believed to be involved in the ring-opening and closing steps.^{40–43} Residues located in the second and third layers from the substrate have been shown to contribute to the catalytic activity.^{5,6} Figure 7 shows the local structure of the multilayer active site of human PGI, with β -mercaptoethanol and a sulfate ion bound (showing the A subunit of the homodimer structure). The catalytic glutamate E358 is shown in teal; coupled residues from the A subunit are colored green; the coupled H389 from the B subunit is shown in gray. Based on PDB structure 1IAT.⁴⁴ Image rendered in Pymol 2.4.1



SCHEME 4 Reaction catalyzed by phosphoglucose isomerase

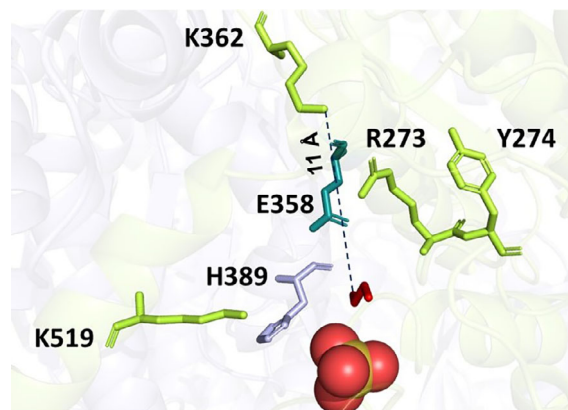


FIGURE 7 Local structure of the multilayer active site of human phosphoglucose isomerase with β -mercaptoethanol and sulfate ion (in red) bound (showing the A subunit of the homodimer structure). The catalytic glutamate E358 is shown in teal; coupled residues from the A subunit are colored green; the coupled H389 from the B subunit is shown in gray. Based on PDB structure 1IAT.⁴⁴ Image rendered in Pymol 2.4.1

environment to 5.9, above that of the two H389s (intrinsic pK_a 4.2), so that the buffer ranges of all four residues are broadened. Although the intrinsic pK_a s of the catalytic glutamates E358 are increased, their pK_a s, calculated with interactions with charged residues included, are heavily downshifted to an estimated value of about -3 , corresponding to strong acidity. This is achieved through interactions with the positively charged residues R273, K362, and K519. K362 is a second-shell residue located behind E358 and R273. Although K362 is positioned 11 Å away from the ligand, it is essential for catalysis; the K362A variant exhibited no detectable activity.^{5,6}

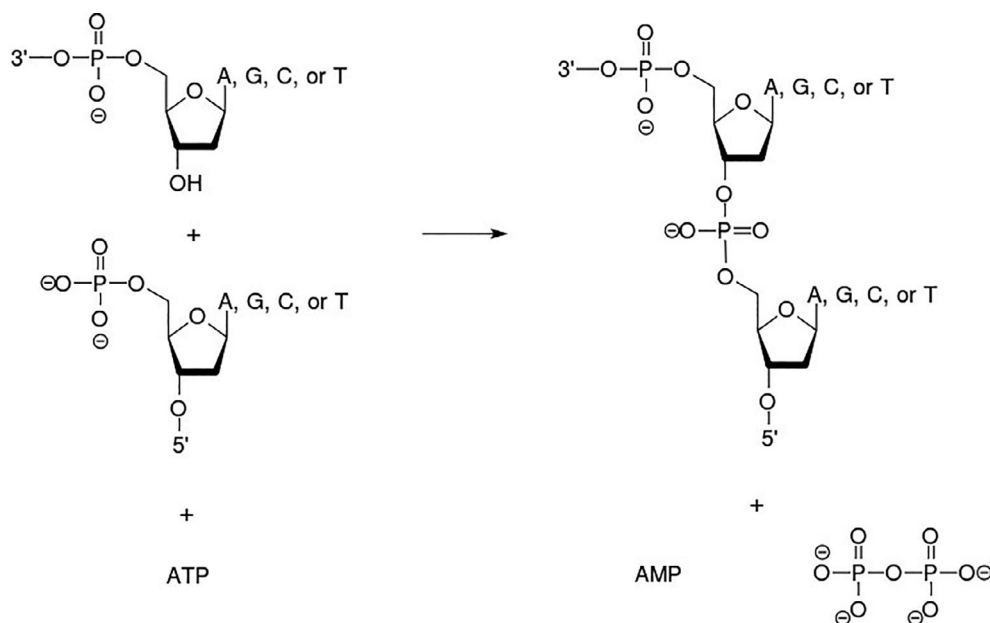
2.2 | Catalytic lysine residues

2.2.1 | DNA ligase

One example of an enzyme with catalytically active lysine residues is ATP-dependent DNA ligase. DNA ligase is important in DNA repair and catalyzes the formation of

TABLE 6 Pairwise electrostatic potential energies ($|E|$, kcal/mol) and intrinsic pK_a s for the five residues most strongly coupled to the catalytic base E358 for the A and B subunits of human phosphoglucose isomerase

Top couplers to E358A: $pK_a(\text{intrinsic}) = 5.9$			Top couplers to E358B: $pK_a(\text{intrinsic}) = 5.9$		
Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$	Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$
R273A	2.3	12.2	R273B	2.3	12.1
H389B	1.2	4.2	H389A	1.2	4.2
K362A	1.2	10.6	K362B	1.2	10.6
Y274A	0.85	12.0	K519B	0.87	9.1
K519A	0.82	9.1	Y274B	0.85	12.0

SCHEME 5 Reaction catalyzed by DNA ligase

phosphodiester linkages between nucleotides (Scheme 5). For DNA ligase from bacteriophage T7 (Uniprot P00969; EC 6.5.1.1; PDB 1A0I),⁴⁶ K34 serves as the nucleophile and K238 and K240 are reported to help to stabilize the transition state.^{47–52} Table 7 and Figure 8 show the coupled partners to K34 that can affect a broadened buffer range (see also Table S12). These include other cation-forming residues, K238, K240 and R55, and the anion-forming residue Y149. These four coupled amino acids have intrinsic pK_a differences with K34 in the range that promotes an expanded buffer range, with small intrinsic pK_a differences for K238, and K240, with a small contribution from R55. Y149 has a higher intrinsic pK_a than K34 by 1.2 pH units, which is within the range of Equation (5), that is, between $|E|/\ln(10)RT - 1$ and $|E|/\ln(10)RT + 1$, in this case between -0.4 and $+1.6$ for $T = 293$ K.

2.2.2 | 1-Aminocyclopropane-1-carboxylate deaminase

The ethylene-sequestering, pyridoxal phosphate (PLP)-dependent enzyme 1-aminocyclopropane-1-carboxylate

TABLE 7 Intrinsic pK_a s for ATP-dependent DNA ligase and energies of interaction with K34

Residue	Intrinsic pK_a	$ E (\text{kcal/mol})$
K34 (nucleophile)	10.1	—
K238	9.9	0.65
K240	10.0	0.87
R55	11.2	0.62
Y149	11.3	0.80

deaminase (ACCD) is found in soil microbes and helps to promote the growth of plants and to protect plants from stresses.⁵³ In *Cyberlindnera saturnus* ACCD (Uniprot Q7M523; EC 3.5.99.7; PDB 1F2D), Lys51 serves as the nucleophile and Tyr269 and Tyr295 are believed to act as proton relays.⁵⁴ The reported lack of activity of the Y295F variant implies that Tyr295 is essential for catalysis and it has been suggested that Tyr295 could also act as a nucleophile.⁵⁵ The residues strongly coupled to Lys51 are: Lys54, Tyr295, Cys199, Glu296, Cys165, and Tyr269 (Table S13). The intrinsic pK_a of K51 is downshifted by

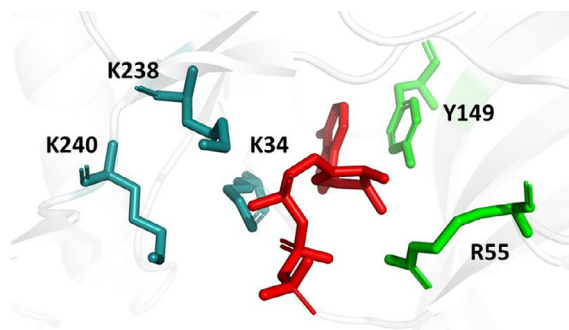


FIGURE 8 Coupled amino acids around the active site of DNA ligase from bacteriophage T7. Catalytic residues K34 (nucleophile) and K238 and K240 (transition state stabilizers) are shown in teal. Coupled residues are shown in green. Ligand is rendered in red. Based on PDB structure 1A0I.⁴⁶ Image rendered in Pymol 2.4.1

the protein environment to the low value of 6.3. The side chain of K54 is located 4 Å away from that of K51; the intrinsic pK_a difference of 1.0 pH units between K51 and K54 contributes to the broadened buffer ranges of both residues. The anion-forming residues Y295, C199, C165, Y269, and C200 have intrinsic pK_a s higher than that of K51 (differences of 1.2, 0.4, 1.3, 0.9, and 1.2, respectively) and within the range to affect broadening of the buffer range, according to Equation (5).

2.2.3 | Dihydrodipicolinate synthase

Dihydrodipicolinate synthase (DHDPS) catalyzes one step in the biosynthesis of lysine, forming 4-hydroxy-tetrahydrodipicolinic acid from L-aspartate- β -semialdehyde and pyruvate. In *E. coli* DHDPS (Uniprot P0A6L2; EC 4.3.3.7; PDB 1DHP), K161 acts as the nucleophile, attacking the keto group of pyruvate and forming a Schiff base.⁵⁶ Y133 is reported to form a hydrogen bond with the carbonyl group of pyruvate, helping to activate it for nucleophilic attack; R138 coordinates the carboxyl group of L-aspartate- β -semialdehyde.⁵⁶ The primary contribution to the broadening of the buffer range of K161 is the interaction with Y133. The potential energy of interaction in units of $\ln(10)RT$ is 2.0, thus the intrinsic pK_a difference of 1.7 (with Y133 higher) is within the range to affect broadening (Table S14).

2.2.4 | 8-oxoguanine DNA-glycosylase

Human 8-oxoguanine DNA-glycosylase (Uniprot O15527; EC 4.2.99.18; PDB 1LWY) catalyzes an important step in DNA repair, excising the modified base 8-oxoguanine from damaged DNA. K249 is the nucleophile, attacking

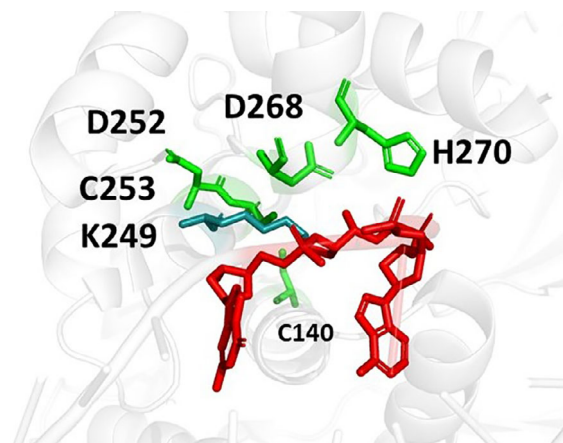
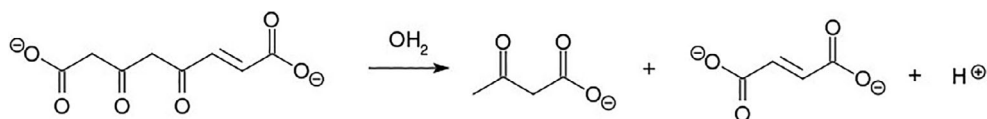


FIGURE 9 Coupled amino acids around the active site of human 8-oxoguanine DNA-glycosylase. The catalytic nucleophile K249 is shown in teal; coupled residues in green. Ligand is shown in red. Based on PDB structure 1LWY.⁵⁹ Image rendered in Pymol 2.4.1

the sugar moiety of 8-oxoguanine DNA, leading to ring opening, release of the 8-oxoguanine, and yielding an abasic site.^{57,58} D268 is reported to stabilize the transition state. K249 is strongly coupled to the anion-forming residue C253 and less strongly to C140. The intrinsic pK_a s of C253 and C140, 9.7 and 10.2, respectively, are higher than the intrinsic pK_a of 9.1 for the nucleophilic K249 and are both well within the range of Equation (5) for expansion of the buffer range (Table S15). C253 is a second-shell residue and in the x-ray crystal structure⁵⁹ it is located behind K249 with respect to the substrate, with the side chains 4.6 Å apart (Figure 9). C140, a third-shell residue situated behind C253 and with weaker potential energy of interaction with K249, makes a smaller contribution to the buffer range.

2.2.5 | O-Phosphoseryl-tRNA: selenocysteinyl-tRNA synthase

The final step in the formation of selenocysteine is catalyzed by O-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS).⁶⁰ SepSecS from *Methanococcus maripaludis* (Uniprot Q6LZM9; EC 2.9.1.2; PDB 2Z67)⁶¹ is dependent on the pyridoxal phosphate (PLP) cofactor and the catalytic K278 is covalently bound to PLP in the initial state of the enzyme as an internal aldimine. Upon binding of substrate, K278 attacks the phosphoserine-bound tRNA to form an external aldimine. K278 serves as both a nucleophile and general acid/base in the reaction mechanism. K278 achieves a broadened buffer range through coupling to another residue in the active site pocket, Y250. The intrinsic pK_a of Y250 is 1.5 pH units



SCHEME 6 Reaction catalyzed by fumarylacetoacetate hydrolase (FAH)

higher than that of K278 and is within the range necessary to affect broadening (Table S16). Y161 and Y257 are also among the top couplers to K278, but their weaker energies of interaction and their intrinsic pK_a differences that exceed the range of Equation (5), suggest that they do not contribute significantly to the broadened buffer range of K278.

2.2.6 | Fumarylacetoacetate hydrolase

A carbon–carbon bond in fumarylacetoacetate is cleaved by fumarylacetoacetate hydrolase (FAH) to yield fumarate and acetoacetate⁶² (Scheme 6). In *Mus musculus* FAH (Uniprot P35505; EC 3.7.1.2; PDB 1HYO), K253 is a general acid/base catalyst.⁶² The strongest coupling partners with K253 are shown in Table 8. The intrinsic pK_a of K253 has been lowered by the protein environment to 8.8. However, strong interactions with five negatively charged neighbors, E199, D233, D126, E201, and D197, has raised the pK_a of the conjugate acid of K253 to an estimated 17, corresponding to strong basicity. The buffer range of K253 is broadened by its strong interaction with Y159 ($\epsilon = 1.4$ in units of $\ln(10)RT$ at $T = 293$ K), which has an intrinsic pK_a 1.5 pH units higher than that of K253, within the range to affect broadening.

2.2.7 | L-amino acid oxidase

The structure of L-amino acid oxidase, a component of the venom from the snake *Calloselasma rhodostoma* (Uniprot P81382; EC 1.4.3.2; PDB 1F8R) that catalyzes the oxidative deamination of L-amino acids, has been reported.⁶³ A FAD co-factor is noncovalently bound to the enzyme. H223 and K326 are reported to serve as the general acids and bases.⁶³ The residues most strongly coupled to the catalytic K326, which has an intrinsic pK_a of 9.1, are shown in Table S17: Y372 and Y356 have intrinsic pK_a s of 10.0 and 10.9, respectively; both have intrinsic pK_a s higher than that of K326 and within the range to affect a broadened buffer range. Both Y372 and Y356 are first-shell residues that interact directly with the substrate; Y372 also makes contact with the FAD co-factor.

TABLE 8 Pairwise energies of interaction (kcal/mol) and intrinsic pK_a s between catalytic K253 and its strongest coupling partners in *Mus musculus* fumarylacetoacetate hydrolase

Top couplers to K253: $pK_a(\text{intrinsic}) = 8.8$		
Residue	$ E (\text{kcal/mol})$	$pK_a(\text{intrinsic})$
E199	3.1	5.3
R237	2.3	10.5
D233	1.9	5.3
Y159	1.8	10.3
D126	1.2	5.0
E201	1.1	5.1
D197	1.0	4.3

3 | DISCUSSION AND CONCLUSIONS

The examples reported here represent all six major EC classifications and multiple different fold types (Table S1). In this diverse set of enzymes, some general patterns emerge in the types of couplings obtained for the catalytic residues.

The catalytic aspartates and glutamates studied here show a pattern of strong coupling to at least one other aspartate or glutamate and often to multiple other carboxylate residues with intrinsic pK_a differences less than or equal to 1 pH unit. Sometimes these catalytic acidic residues are also coupled to histidines, such that the intrinsic pK_a of the acidic residue is higher than that of the histidine. This is the case for the catalytic E358 in phosphoglucose isomerase, for instance; a histidine with an unusually downshifted intrinsic pK_a is the primary contributor to an expanded buffer range. These trends are summarized in Table 9.

For the catalytic lysine residues studied here, all are strongly coupled to tyrosine or cysteine residues, wherein the intrinsic pK_a of the anion-forming residue is higher than that of the lysine. Some catalytic lysines are also coupled to other lysines with intrinsic pK_a differences within 1 pH unit. These features are summarized in Table 10.

For the glycoside hydrolases shown in Table 1, there is a distinct pattern wherein both catalytic glutamates have very downshifted pH profiles, corresponding to low

values for μ_1 compared to the other glutamates, but the nucleophilic glutamate has the lowest value for μ_1 . To have both protonation states in significant population at neutral pH, interactions with nearby residues create a wide buffer range on the high pH side of the titration curve, and this is reflected in the high, positive values for μ_3 for the nucleophile. This pattern of low μ_1 and high, positive μ_3 is also observed for the catalytic nucleophile D8 in haloacid dehalogenase (Table S18). However, the

reported catalytic nucleophile in DNase, D41 (Table S18) does not show this pattern, suggesting that there is more than one way to make an acidic residue into a nucleophile. Direct experimental verification of nucleophilicity via suicide inhibition would be helpful.

For the lysine nucleophiles, this pattern of low μ_1 and high, positive μ_3 is found for K51 of 1-amino-cyclopropane-1-carboxylate deaminase (ACCD). A titration curve shifted to higher pH, corresponding to a

TABLE 9 Summary of the most strongly coupled amino acids that contribute to an expanded buffer range for catalytically active aspartates and glutamates

Enzyme	PDB	Catalytic residue	Role	Coupled residue	ΔpK_a intrinsic	E (kcal/mol)
β -Glucosidase	1OIF	E166	Acid/base	E351	+0.6	1.9
				E405	+0.6	0.95
		E351	Nucleophile	E166	−0.6	1.9
				E405	0.0	1.0
DNase	1Q91	D41	Nucleophile	D175	+0.3	3.0
				D176	+0.1	1.7
				D43	0.0	1.2
		D43	Acid/base	D41	0.0	1.2
				D176	+0.1	0.97
Haloacid dehalogenase	1QQ5	D8	Nucleophile	D176	+0.6	2.7
		D176	Acid/base	D8	−0.6	2.7
Ornithine cyclodeaminase	1X7D	E56	Acid/base	E256	+0.1	1.3
				D228	−0.3	1.3
		D228	Acid/base	E56	+0.3	1.3
				E256	+0.4	0.94
HGPRT	1BZY	D137	Acid/base	E133	−0.6	0.76
				D184	−0.7	0.75
Asp ammonia ligase	12AS	D46	Acid/base	D118	+0.7	0.94
Fucose isomerase	1FUI	E337	Acid/base	H528	−0.7	1.5
				D361	+0.4	1.3
				E316	−0.9	1.2
				D313	−0.8	1.1
				D339	−0.5	0.83
				D361	−0.4	1.3
				H528	−1.1	1.2
				D339	−0.9	0.56
Pyrogallol hydroxytransferase	4V4E	D174	Acid/base	E177	+0.3	1.6
				E818	+0.5	1.3
Phosphoglucose isomerase	1IAT	E358A	Acid/base	H389B	−1.7	1.2
				D511A	−0.9	0.70
				D356A	−0.6	0.65

Note: The difference in intrinsic pK_a , ΔpK_a , is expressed as [pK_a (intrinsic, coupler) − pK_a (intrinsic, catalytic)], thus a positive value means that the coupled partner has a higher pK_a than the catalytic residue. Histidines with a negative ΔpK_a (intrinsic) have a lower intrinsic pK_a than the catalytic acid.

TABLE 10 Summary of the most strongly coupled amino acids that contribute to an expanded buffer range for catalytically active lysines

Enzyme	PDB	Catalytic residue	Role	Coupled residue	ΔpK_a	E (kcal/mol)
DNA ligase	1A0I	K34	Nucleophile	K238	−0.2	0.65
				K240	−0.1	0.87
				Y149	+1.2	0.80
1-Aminocyclopropane-1-carboxylate deaminase (ACCD)	1F2D	K51	Nucleophile	K54	+1.0	2.7
				Y295	+2.2	1.7
				C199	+1.4	1.3
				C165	+2.3	1.2
				Y269	+1.9	1.2
				C200	+2.2	1.0
Dihydrodipicolinate synthase	1DHP	K161	Nucleophile	Y133	+1.7	2.6
8-Oxoguanine DNA-glycosylase	1LWY	K249	Nucleophile	C253	+0.6	1.8
				C140	+1.1	0.5
O-phosphoseryl-tRNA:Selenocysteinyl-tRNA synthase	2Z67	K278	Nucleophile	Y250	+1.5	1.1
				Y161	+3.0	0.49
				Y257	+2.1	0.44
Fumarylacetoacetate hydrolase	1HYO	K253	Acid/base	Y159	+1.5	1.8
L-amino acid oxidase	1F8R	K326	Acid/base	Y372	+0.9	1.7
				Y356	+1.8	0.91

Note: The difference in intrinsic pK_a s, ΔpK_a , is expressed as [$pK_a(\text{intrinsic, coupler}) - pK_a(\text{intrinsic, catalytic})$], thus a positive value means that the coupled partner has a higher pK_a than the catalytic residue. Anion-forming residues such as tyrosine and cysteine with a positive $\Delta pK_a(\text{intrinsic})$ have a higher intrinsic pK_a than the catalytic lysine.

larger, positive μ_1 , with a long tail on the low pH side, corresponding to a negative μ_3 , is another way to achieve significant population of the deprotonated state at neutrality; this is observed in the theoretical titration behavior of K161 of dihydrodipicolinate synthase (DHDPs, Tables S14 and S19) and of K249 of 8-oxoguanine DNA-glycosylase (Tables S15 and S19).

The broad buffer ranges of catalytic aspartates, glutamates, and lysines, achieved through specific types of electrostatic interactions discussed here, are an important factor in the catalytic potency of these residues. The broad buffer range enables enhanced acidity/basicity combined with significant population of both protonation states at the pH at which the enzyme operates, which for most enzymes of most species is near neutrality. The coupled partners that help to give activity to the catalytic residues may be first-shell residues that interact directly with the substrate molecule or may be located in the second or third shell from the substrate.

The present results provide important understanding of some of the interactions between residues that transform functional groups that are weak acids or bases when free in solution into powerful acids, bases, or nucleophiles inside the protein structure. Understanding the

interactions that promote catalytic activity can help to identify residue roles and thus could provide some information about the biochemical function of Structural Genomics protein structures of unknown function. These concepts may also provide valuable information for efforts to design novel enzymes.

4 | MATERIALS AND METHODS

Three-dimensional structures for proteins were downloaded from the Protein Data Bank⁶⁴ and preprocessed using YASARA⁶⁵ to insert any missing atoms. For aspartate ammonia ligase, the sequence of the double-mutant C51A/C315A structure (PDB 12AS)³⁶ was edited back to the wild-type sequence and the wild-type structure was built using the mutant structure as the template and the homology modeling feature in YASARA.⁶⁶ Electrical potential functions were calculated by a linear Poisson–Boltzmann method^{9,10,67,68} at a temperature of 293 K and theoretical titration curves were calculated by the hybrid method,⁶⁹ as described previously.^{23,70} Information about catalytic residues and their mechanistic roles was obtained from the cited literature and from the

Mechanism and Catalytic Site Atlas (M-CSA).⁷¹ Central moments were obtained numerically as described previously.^{3,4} Computed central moments for catalytic aspartate and glutamate residues are shown in Tables 1 and S17. Computed central moments for catalytic lysine residues are shown in Table S19.

ACKNOWLEDGMENTS

This work is supported by NSF CHE-1905214 and by a Fulbright Research Grant 2221103 (Mary Jo Ondrechen). The authors thank Professor Dagmar Ringe for valuable discussions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Suhasini M. Iyengar: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); writing – review and editing (equal). **Kelly K. Barnsley:** Data curation (equal); investigation (equal); visualization (equal); writing – review and editing (equal). **Rholee Xu:** Investigation (equal); methodology (equal); software (equal); writing – review and editing (supporting). **Aleksandr Prystupa:** Data curation (equal); investigation (equal); writing – review and editing (equal). **Mary Jo Ondrechen:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); writing – original draft (equal); writing – review and editing (equal).

ORCID

Mary Jo Ondrechen  <https://orcid.org/0000-0003-2456-4313>

REFERENCES

- Agarwal PK. A biophysical perspective on enzyme catalysis. *Biochemistry*. 2019;58:438–449.
- Mazmanian K, Sargsyan K, Lim C. How the local environment of functional sites regulates protein function. *J Am Chem Soc*. 2020;142:9861–9871.
- Wei Y, Ko J, Murga L, Ondrechen MJ. Selective prediction of interaction sites in protein structures with THEMATICS. *BMC Bioinformatics*. 2007;8:119.
- Ko J, Murga LF, André P, et al. Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins*. 2005;59:183–195.
- Somarowthu S, Brodtkin HR, D'Aquino JA, Ringe D, Ondrechen MJ, Beuning PJ. A tale of two isomerases: Compact versus extended active sites in ketosteroid isomerase and phosphoglucose isomerase. *Biochemistry*. 2011;50:9283–9295.
- Brodtkin HR, DeLateur NA, Somarowthu S, et al. Prediction of distal residue participation in enzyme catalysis. *Protein Sci*. 2015;24:762–778.
- Brodtkin HR, Novak WR, Milne AC, et al. Evidence of the participation of remote residues in the catalytic activity of co-type nitrile hydratase from *Pseudomonas putida*. *Biochemistry*. 2011;50:4923–4935.
- Ringe D, Wei Y, Boino KR, Ondrechen MJ. Protein structure to function: Insights from computation. *Cell Mol Life Sci*. 2004; 61:387–392.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*. 2001;98:10037–10041.
- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004; 32:W665–W667.
- Ondrechen MJ, Clifton JG, Ringe D. THEMATICS: A simple computational predictor of enzyme function from structure. *Proc Natl Acad Sci U S A*. 2001;98:12473–12478.
- Di Cera E, Gill SJ, Wyman J. Binding capacity: Cooperativity and buffering in biopolymers. *Proc Natl Acad Sci U S A*. 1988; 85:449–452.
- Coulther TA, Ko J, Ondrechen MJ. Amino acid interactions that facilitate enzyme catalysis. *J Chem Phys*. 2021;154:195101.
- Ullmann GM. Relations between protonation constants and titration curves in polyprotic acids: A critical view. *J Phys Chem B*. 2003;107:1263–1271.
- Koumanov A, Rüterjans H, Karshikoff A. Continuum electrostatic analysis of irregular ionization and proton allocation in proteins. *Proteins*. 2002;46:85–96.
- Mackenzie LF, Brooke GS, Cutfield JF, Sullivan PA, Withers SG. Identification of Glu-330 as the catalytic nucleophile of *Candida albicans* exo-beta-(1,3)-glucanase. *J Biol Chem*. 1997;272:3161–3167.
- White A, Tull D, Johns K, Withers SG, Rose DR. Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase. *Nat Struct Mol Biol*. 1996;3:149–154.
- Ryttersgaard C, Leggio LL, Coutinho PM, Henrissat B, Larsen S. *Aspergillus aculeatus* beta-1,4-galactanase: Substrate recognition and relations to other glycoside hydrolases in clan GH-A. *Biochemistry*. 2002;41:15135–15143.
- Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry*. 1996;35:10648–10660.
- Zechel DL, Boraston AB, Gloster T, Boraston CM, Macdonald JM, Tilbrook DMG, Stick RV, Davies GJ. Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to β -glucosidases. *Journal of the American Chemical Society*. 2003;125(47):14313–14323. <https://doi.org/10.1021/ja036833h>
- Juers DH, Heightman TD, Vasella A, McCarter JD, Mackenzie L, Withers SG, Matthews BW. A structural view of the action of *Escherichia coli* (lacZ) β -galactosidase. *Biochemistry*. 2001;40(49):14781–14794. <https://doi.org/10.1021/bi011727i>
- Cutfield SM, Davies GJ, Murshudov G, et al. The structure of the exo-beta-(1,3)-glucanase from *Candida albicans* in native

- and bound forms: Relationship between a pocket and groove in family 5 glycosyl hydrolases. *J Mol Biol.* 1999;294:771–783.
23. Tong W, Wei Y, Murga LF, Ondrechen MJ, Williams RJ. Partial order optimum likelihood (POOL): Maximum likelihood prediction of protein active site residues using 3D structure and sequence properties. *PLoS Comput Biol.* 2009;5:e1000266.
 24. Marana SR, Mendonça LMF, Andrade EHP, Terra WR, Ferreira C. The role of residues R97 and Y331 in modulating the pH optimum of an insect β -glycosidase of family 1. *Eur J Biochem.* 2003;270:4866–4875.
 25. Sondergaard CR, McIntosh LP, Pollastri G, Nielsen JE. Determination of electrostatic interaction energies and protonation state populations in enzyme active sites. *J Mol Biol.* 2008;376:269–287.
 26. McIntosh LP, Hand G, Johnson PE, et al. The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A ¹³C-NMR study of *Bacillus circulans* xylanase. *Biochemistry.* 1996;35:9958–9966.
 27. Ridder IS, Rozeboom HJ, Kalk KH, Dijkstra BW. Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. *J Biol Chem.* 1999;274:30672–30678.
 28. Rinaldo-Matthis A, Rampazzo C, Reichard P, Bianchi V, Nordlund P. Crystal structure of a human mitochondrial deoxyribonucleotidase. *Nat Struct Biol.* 2002;9:779–787.
 29. Seifried A, Schultz J, Gohla A. Human HAD phosphatases: Structure, mechanism, and roles in health and disease. *FEBS J.* 2013;280:549–571.
 30. Nardi-Dei V, Kurihara T, Park C, Esaki N, Soda K. Bacterial DL-2-haloacid dehalogenase from pseudomonas sp. strain 113: Gene cloning and structural comparison with D- and L-2-haloacid dehalogenases. *J Bacteriol.* 1997;179:4232–4238.
 31. Li YF, Hata Y, Fujii T, et al. Crystal structures of reaction intermediates of L-2-haloacid dehalogenase and implications for the reaction mechanism. *J Biol Chem.* 1998;273:15035–15044.
 32. Goodman JL, Wang S, Alam S, Ruzicka FJ, Frey PA, Wedekind JE. Ornithine cyclodeaminase: structure, mechanism of action, and implications for the μ -crystallin family. *Biochemistry.* 2004;43(44):13883–13891. <https://doi.org/10.1021/bi048207i>
 33. Shi W, Li CM, Tyler PC, et al. The 2.0 Å structure of human hypoxanthine-guanine phosphoribosyltransferase in complex with a transition-state analog inhibitor. *Nat Struct Biol.* 1999;6:588–593.
 34. Héroux A, White EL, Ross LJ, Davis RL, Borhani DW. Crystal structure of toxoplasma gondii hypoxanthine-guanine phosphoribosyltransferase with xmp, pyrophosphate, and two Mg²⁺ ions bound: insights into the catalytic mechanism. *Biochemistry.* 1999;38(44):14495–14506. <https://doi.org/10.1021/bi990508i>
 35. Koizumi M, Hiratake J, Nakatsu T, Kato H, Oda J. A potent transition-state analogue inhibitor of escherichia coli asparagine synthetase A. *Journal of the American Chemical Society.* 1999;121(24):5799–5800. <https://doi.org/10.1021/ja990851a>
 36. Nakatsu T, Kato H, Oda J. Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. *Nat Struct Biol.* 1998;5:15–19.
 37. Seemann JE, Schulz GE. Structure and mechanism of L-fucose isomerase from *Escherichia coli*. *J Mol Biol.* 1997;273:256–268.
 38. Messerschmidt A, Niessen H, Abt D, Einsle O, Schink B, Kroneck PMH. Crystal structure of pyrogallol-phloroglucinol transhydroxylase, an Mo enzyme capable of intermolecular hydroxyl transfer between phenols. *Proc Natl Acad Sci U S A.* 2004;101:11571–11576.
 39. Paizs C, Bartlewski-Hof U, Rétey J. Investigation of the mechanism of action of pyrogallol-phloroglucinol transhydroxylase by using putative intermediates. *Chemistry - A European Journal.* 2007;13(10):2805–2811. <https://doi.org/10.1002/chem.200601053>
 40. Jeffery CJ, Hardré R, Salmon L. Crystal Structure of Rabbit Phosphoglucose Isomerase Complexed with 5-Phospho-d-Arabinonate Identifies the Role of Glu357 in Catalysis. *Biochemistry.* 2001;40(6):1560–1566. <https://doi.org/10.1021/bi0018483>
 41. Lee JH, Chang KZ, Patel V, Jeffery CJ. Crystal structure of rabbit phosphoglucose isomerase complexed with its substrate d-fructose 6-phosphate. *Biochemistry.* 2001;40:7799–7805.
 42. Solomons JTG, Zimmerly EM, Burns S, et al. The crystal structure of mouse phosphoglucose isomerase at 1.6 Å resolution and its complex with glucose 6-phosphate reveals the catalytic mechanism of sugar ring opening. *J Mol Biol.* 2004;342:847–860.
 43. Davies C, Muirhead H, Chirgwin J. The structure of human phosphoglucose isomerase complexed with a transition-state analogue. *Acta Crystallogr D Biol Crystallogr.* 2003;59:1111–1113.
 44. Read J, Pearce J, Li X, Muirhead H, Chirgwin J, Davies C. The crystal structure of human phosphoglucose isomerase at 1.6 Å resolution: Implications for catalytic mechanism, cytokine activity and haemolytic anaemia. *J Mol Biol.* 2001;309:447–463.
 45. Ahmad L, Plancqueel S, Lazar N, et al. Novel N-substituted 5-phosphate-d-arabinonamide derivatives as strong inhibitors of phosphoglucose isomerases: Synthesis, structure-activity relationship and crystallographic studies. *Bioorg Chem.* 2020;102:104048.
 46. Doherty AJ, Ashford SR, Subramanya HS, Wigley DB. Bacteriophage T7 DNA ligase: Overexpression, purification, crystallization, and characterization. *J Biol Chem.* 1996;271:11083–11089.
 47. Odell M, Sriskanda V, Shuman S, Nikolov DB. Crystal structure of eukaryotic DNA ligase—adenylate illuminates the mechanism of nick sensing and strand joining. *Mol Cell.* 2000;6:1183–1193.
 48. Cherepanov AV, de Vries S. Kinetic mechanism of the Mg²⁺–dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases. *J Biol Chem.* 2002;277:1695–1704.
 49. Luo J, Barany F. Identification of essential residues in *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* 1996;24:3079–3085.
 50. Shuman S, Ru XM. Mutational analysis of vaccinia DNA ligase defines residues essential for covalent catalysis. *Virology.* 1995;211:73–83.
 51. Sriskanda V, Shuman S. Mutational analysis of Chlorella virus DNA ligase: Catalytic roles of domain I and motif VI. *Nucleic Acids Res.* 1998;26:4618–4625.
 52. Tomkinson AE, Totty NF, Ginsburg M, Lindahl T. Location of the active site for enzyme-adenylate formation in DNA ligases. *Proc Natl Acad Sci U S A.* 1991;88:400–404.

53. Polko JK, Kieber JJ. 1-Aminocyclopropane 1-carboxylic acid and its emerging role as an ethylene-independent growth regulator. *Front Plant Sci.* 2019;10:1602.
54. Hontzeas N, Hontzeas CE, Glick BR. Reaction mechanisms of the bacterial enzyme 1-aminocyclopropane-1-carboxylate deaminase. *Biotechnol Adv.* 2006;24:420–426.
55. Karthikeyan S, Zhao Z, Kao C-I, et al. Structural analysis of 1-aminocyclopropane-1-carboxylate deaminase: Observation of an aminyl intermediate and identification of Tyr 294 as the active-site nucleophile. *Angew Chem Intl Ed.* 2004;43:3425–3429.
56. Blickling S, Renner C, Laber B, Pohlenz H-D, Holak TA, Huber R. Reaction mechanism of *Escherichia coli* dihydroadipic acid synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry.* 1997;36:24–33.
57. Chung SJ, Verdine GL. Structures of end products resulting from lesion processing by a DNA glycosylase/lyase. *Chem Biol.* 2004;11:1643–1649.
58. Kellie JL, Wetmore SD. Mechanistic and conformational flexibility of the covalent linkage formed during β -lyase activity on an AP-site: Application to hOgg1. *J Phys Chem B.* 2012;116:10786–10797.
59. Fromme JC, Bruner SD, Yang W, Karplus M, Verdine GL. Product-assisted catalysis in base-excision DNA repair. *Nat Struct Biol.* 2003;10:204–211.
60. Palioura S, Sherrer RL, Steitz TA, Söll D, Simonovic M. The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science.* 2009;325:321–325.
61. Araiso Y, Palioura S, Ishitani R, et al. Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res.* 2008;36:1187–1199.
62. Bateman RL, Bhanumoorthy P, Witte JF, McClard RW, Grompe M, Timm DE. Mechanistic inferences from the crystal structure of fumarylacetoacetate hydrolase with a bound phosphorus-based inhibitor. *J Biol Chem.* 2001;276:15284–15291.
63. Moustafa IM, Foster S, Lyubimov AY, Vrielink A. Crystal structure of LAAO from *Calloselasma rhodostoma* with an l-phenylalanine substrate: Insights into structure and mechanism. *J Mol Biol.* 2006;364:991–1002.
64. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28:235–242.
65. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins.* 2004;57:678–683.
66. Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins.* 2009;77:114–122.
67. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. *J Mol Biol.* 1994;238:415–436.
68. Shen T, Wong CF, McCammon JA. Brownian dynamics simulation of helix-capping motifs. *Biopolymers.* 2003;70:252–259.
69. Gilson MK. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins.* 1993;15:266–282.
70. Somarowthu S, Yang H, Hildebrand DGC, Ondrechen MJ. High-performance prediction of functional residues in proteins with machine learning and computed input features. *Biopolymers.* 2011;95:390–400.
71. Ribeiro AJM, Holliday GL, Furnham N, Tyzack JD, Ferris K, Thornton JM. Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* 2018;46:D618–D623.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Iyengar SM, Barnsley KK, Xu R, Prystupa A, Ondrechen MJ. Electrostatic fingerprints of catalytically active amino acids in enzymes. *Protein Science.* 2022; 31(5):e4291. <https://doi.org/10.1002/pro.4291>