


FULL-LENGTH PAPER

Analysis of electrostatic coupling throughout the laboratory evolution of a designed retroaldolase

Timothy A. Coulther^{1,3} | Moritz Pott² | Cathleen Zeymer^{2,4} | Donald Hilvert² | Mary Jo Ondrechen¹ 

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

²Laboratory of Organic Chemistry, ETH Zürich, Zürich, Switzerland

³Genome Center, University of California, Davis, California

⁴Department of Chemistry, Technische Universität München, Garching, Germany

Correspondence

Mary Jo Ondrechen, Department of Chemistry and Chemical Biology, ISEC 129, Northeastern University, Boston, MA 02115.

Email: mjo@neu.edu

Present address

Timothy A Coulther, Genome Center, University of California – Davis, Davis, CA 95616 USA.

Cathleen Zeymer, Department of Chemistry, Technische Universität München, D-85748 Garching, Germany.

Funding information

Marie Skłodowska-Curie Individual Fellowship, Grant/Award Number: TIMEnzyme(CZ); Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; National Institute of Justice, Grant/Award Number: Predoctoral Fellowship (TAC); Swiss National Science Foundation; ETH Zurich; National Science Foundation, Grant/Award Numbers: CHE-1905214, MCB-1517290

Abstract

The roles of local interactions in the laboratory evolution of a highly active, computationally designed retroaldolase (RA) are examined. Partial Order Optimum Likelihood (POOL) is used to identify catalytically important amino acid interactions in several RA95 enzyme variants. The series RA95.5, RA95.5–5, RA95.5–8, and RA95.5–8F, representing progress along an evolutionary trajectory with increasing activity, is examined. Computed measures of coupling between charged states of residues show that, as evolution proceeds and higher activities are achieved, electrostatic coupling between the biochemically active amino acids and other residues is increased. In silico residue scanning suggests multiple coupling partners for the catalytic lysine K83. The effects of two predicted partners, Y51 and E85, are tested using site-directed mutagenesis and kinetic analysis of the variants Y51F and E85Q. The Y51F variants show decreases in k_{cat} relative to wild type, with the greatest losses observed for the more evolved constructs; they also exhibit significant decreases in $k_{\text{cat}}/K_{\text{M}}$ across the series. Only modest decreases in $k_{\text{cat}}/K_{\text{M}}$ are observed for the E85Q variants with little effect on k_{cat} . Computed metrics of the degree of coupling between protonation states rise significantly as evolution proceeds and catalytic turnover rate increases. Specifically, the charge state of the catalytic lysine K83 becomes more strongly coupled to those of other amino acids as the enzyme evolves to a better catalyst.

KEYWORDS

enzyme design, kinetics, partial order optimum likelihood, protein electrostatics, protonation states, retroaldolases, site-directed mutagenesis

Abbreviations: POOL, partial order optimum likelihood; RA95, designed retroaldolase, series 95; THEMATICS, theoretical microscopic titration curve shapes; μ_n , n^{th} central moment.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Enzymes, with their exquisite selectivity and catalytic power, have tremendous potential for use in the chemical industry and for diverse applications, such as bioremediation and biofuel production. Indeed enzymes have gained some use in industrial chemical processes; for example nitrile hydratase, which catalyzes the conversion of nitriles to amides, is used in the production of acrylamide on the kiloton scale yearly, at lower cost than conventionally catalyzed processes.^{1,2} However, for many reactions involved in industrial chemical processes, a natural enzyme may not exist. The ability to design enzymes that can perform specific, unnatural reactions can open the door to greener chemical industry, with less energy consumption and fewer unwanted by-products than thermal processes or processes that use more conventional catalysts. This requires expanded knowledge of the fundamental mechanisms by which both natural and designed enzymes work.

Very recently, Mazmanian, Sargsyan and Lim³ reviewed how the environment of enzyme active sites, namely local interactions, the presence of solvent molecules, and conformational flexibility, control enzymatic function. Here we focus on the role of local interactions^{4,5} between catalytic residues and nearby residues on the activity of a series of designed retroaldolases of the RA95 family, which has been optimized by laboratory evolution.^{6–9}

Design of novel enzymes typically involves a new, computationally designed active site built into a natural protein scaffold. While this can be significantly more difficult than engineering natural enzymes for altered specificities, it opens up the functional space of enzymes to catalyze reactions that do not have a natural counterpart. Unfortunately, low activity of initial designs plagues most engineering attempts.

To increase the activity of the designed enzymes, directed evolution is employed; it relies on the screening of many enzyme variants, often with mutations introduced randomly across the sequence. Beneficial mutations are slowly accumulated and combined to produce better enzymes and the process can require many iterations. While this strategy can produce enzymes with activities orders of magnitude greater than the starting design, it is time- and resource-intensive. To date such experiments have taken years to produce enzymes with useful levels of activity. Advances have been made to help directed evolution, such as the use of microfluidics to miniaturize the process and increase throughput,^{9–11} but improved initial designs could reduce substantially the time required to achieve effective activity.

A previously designed retroaldolase, RA95, has been the focus of much research as it has been evolved over

several years and its activity increased over five orders of magnitude from its initial design.^{6,8,9} Kinetics data have been reported for the designs as they evolve from low activity to high activity, including data for multiple steps in between. Designed enzymes from this series were chosen for the present study because measured rate constants span a wide range of values and because crystal structures are available for multiple steps along the evolutionary path. Our computational tools allow us to understand better the sources of increased activity in the RA95 series and ultimately to help to establish the basic concepts for increased activity of designed enzymes. We have identified computed features that correlate with the increase in activity and have been able to attribute activity gains to specific residues in some instances. Using our computational tools such as THEMATICS^{12,13} and POOL,^{14,15} we have identified key mutations during evolution that contribute to increased activity along the evolutionary trajectory. This has also enabled the discovery of key metrics that mirror the activity differences between the enzymes, suggesting these values might be used to guide enzyme designs or to screen candidates computationally. These metrics are based on computed electrostatic and chemical properties of the individual amino acid residues in the protein structure.

2 | RESULTS

2.1 | Predicting chemical properties of protein active sites with POOL

Predictions by POOL^{14,15} of biochemically active residues are based on metrics obtained from computed proton occupation functions $C(\text{pH})$ for each site with proton transfer capability, which are in turn based on the computed electrical potential function 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 0 at high pH, is a sigmoidal function with a sharp fall-off at pH values near the $\text{p}K_{\text{a}0}$. However, for residues active in catalysis and/or ligand recognition, C has an anomalous shape, arising from strong coupling to proton transfer equilibria on nearby residues, much like the behavior observed for polyprotic acids. To quantify the deviation from typical behavior, the $C(\text{pH})$ functions are expressed in terms of a first derivative function f , defined as:

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

The negative first derivative of C , f , is essentially a proton binding capacity,¹⁶ a measure of the change in concentration of a bound species per unit change in its

chemical potential. For a typical titratable residue, the f function may be well approximated by a single Gaussian distribution function, with a peak at or near the pH equal to the pK_a . In contrast, for titratable residues involved in catalysis or ligand binding, the f function deviates significantly from Gaussian form and may be represented as a superposition of Gaussian functions. The area under the f function is always unity because of the 1 to 0 range of the C functions and f is nonnegative; therefore $f(\text{pH})$ resembles a probability density function, the shape of which can be characterized by summary statistics,¹⁷ such as the n^{th} central moments μ_n , defined as:

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

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

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

where the integrals in Equations (2) and (3) are over all real numbers ($-\infty$ to $+\infty$).¹⁷ We have established that the third and fourth central moments, μ_3 and μ_4 of Equation (2), are excellent predictors of the residues active in catalysis and/or ligand recognition for natural enzymes.^{17,18} The third and fourth central moments may be calculated for each titratable residue, using the electrical potential function calculated from the three-dimensional structure of the protein. These features are indicators of the degree of coupling between the proton binding reaction of a given titratable residue and those of other residues; for biochemically active residues in natural enzymes, these couplings are strong.^{12,18–21}

There is considerable evidence that electrostatic effects, including contributions from distal residues, are important in enzyme catalysis. A 1991 density functional theory study was reported by Bajorath et al. of free dihydrofolate (DHF) and of DHF bound to dihydrofolate reductase. They showed that when DHF is placed in the electrical potential of the enzyme, the π orbitals of the reactive C=N bond of DHF are polarized, such that the π bond begins to split into two lobes that resemble atomic 2p orbitals on the C and N atoms. This facilitates reduction wherein σ bonds are formed with hydrogen 1 s orbitals; the polarizing potential was reported to arise significantly from both nearby and remote residues.²² Electric fields in protein binding sites can be substantial. Boxer et al.²³ used vibrational Stark spectroscopy to measure local electric fields in protein binding sites and reported electric fields on the order of 100 MV/cm!

Experimental evidence of significant coupling by distal residues has been reported by us.^{2,5,24–27} For all of these cases, the important couplings of the specific distal residues were predicted computationally. Although there is clear evidence of the importance of electrostatic effects and coupling of distal residues in enzyme catalysis, these effects are often not considered, including in the context of enzyme design.

Directed evolution often utilizes random mutagenesis to identify variants with a new function or higher activity, as was done in the RA95 family. It is often however not known exactly what causes the increased activity, specifically what residues are responsible or how they contribute to increased activity. The answers are generally not obvious. For instance, the introduction of a second lysine (K83) in the evolved variants of the RA95 series was originally believed to function by helping to modulate the pK_a of the originally designed nucleophilic lysine (K210) as its function. However, x-ray crystallography showed that K83 had become an active nucleophile itself. Understanding how and which mutations are beneficial is important to the development of guiding principles for enzyme engineering.

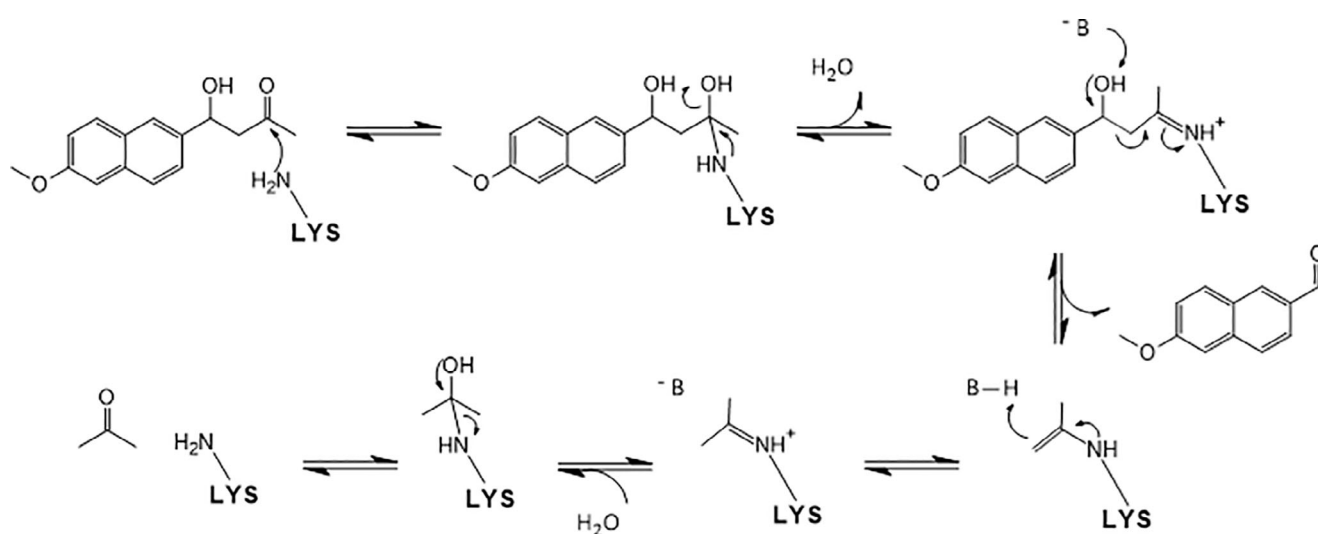
3 | POOL PREDICTIONS AND DIRECTED EVOLUTION RETROSPECTIVE

For three constructs, RA95.5, RA95.5–5, and RA95.5–8, five residues have POOL scores higher than the average score; for RA95.5–8F, seven residues fall above this threshold. The amino acids ranked in the top 12 are shown in Table 1; residues with POOL scores below the average score are shown in italics. For RA95.5, RA95.5–5, and RA95.5–8 both active site lysines, K83 and K210, are predicted to be important for biochemical activity. K210 was the original catalytic residue introduced by design. Over the course of evolution, K83 was introduced and activity gradually shifted from K210 to K83. Thus, the original design only had K210; RA95.5, RA95.5–5, and RA95.5–8 have both. In RA95.5–8F, evolution replaced K210 with leucine.

The catalytic mechanism involves attack of the reactive lysine (K210 and/or K83) on the carbonyl group of the substrate to form a tetrahedral intermediate, which loses water to give a Schiff base adduct (Figure 1). The tyrosine residue Y51 is in contact with K83, located slightly behind K83 with respect to the substrate in RA95.5⁸ but moves closer to the substrate in later constructs. It is ranked highly by POOL for all four constructs. Y51 has been proposed as a likely Brønsted acid/base,⁹ although a computational study has suggested

TABLE 1 The top 12 POOL-predicted biochemically active residues in rank order for members of the RA95 series. Residues in *italics* have POOL scores below the average score for each construct

POOL rank	RA95.5	RA95.5–5	RA95.5–8	RA95.5–8F
PDB	4A2S	4A2R	Model	6TFA
1st	K83	K83	K83	Y51
2nd	K210	Y51	Y51	K83
3rd	Y51	K210	K210	Y180
4th	D111	D111	D111	D212
5th	E85	E85	E85	D111
6th	<i>S233</i>	<i>N110</i>	<i>R54</i>	D61
7th	<i>R19</i>	<i>L231</i>	Y52	E85
8th	<i>S110</i>	<i>R19</i>	<i>N110</i>	<i>E135</i>
9th	<i>F112</i>	<i>K135</i>	<i>S233</i>	<i>L210</i>
10th	<i>D165</i>	<i>L131</i>	<i>L131</i>	<i>R19</i>
11th	<i>L131</i>	<i>F112</i>	<i>F89</i>	<i>N110</i>
12th	<i>I232</i>	Y52	<i>R19</i>	<i>S233</i>

**FIGURE 1** Reaction scheme for the retroaldolase

that Y180 may assume this role in the highly evolved RA95.5–8F.²⁸ Note that Y180, which emerged late in the evolutionary trajectory and is part of a hydrogen-bonded network consisting of Y51, K83, and N110, is ranked third in RA95.5–8F; the earlier constructs all have phenylalanine in this position.

Located near the entrance to the binding pocket, along the rim of the beta barrel, E85 is predicted to be active in all four variants; no role in catalysis has been assigned previously to this residue. The acid side chain of E85 forms hydrogen-bonds with the backbone N-H groups of K87 and Y88, contributing to fold stability. D111, which is situated behind E85, is also predicted for

all four constructs. D111 is conserved throughout the series and is critical to structural integrity; it forms a salt bridge with the conserved R19, helping to position the N-terminal helix–loop, and has a charge–dipole interaction with the alcohol group on the side chain of T84. Two additional aspartate residues, D212 and D61, are predicted for RA95.5–8F. D61 and D212 are located on the rim of the beta barrel at the entrance to the binding pocket, on the side opposite E85 and D111; these acidic residues at the entrance to the binding pocket may help to orient the substrate as it enters the pocket.

Residue N110, which provided a five-fold boost in activity when first introduced by directed evolution,

ranks in the top 12 for RA95.5–5, RA95.5–8, and RA95.5–8F, but has POOL scores below the average score, a somewhat arbitrary cutoff. Note that the original serine at this position ranks eighth in the early construct RA95.5. N110 is located within hydrogen-bonding distance of K83 and its polar side chain may contribute to the catalytic power of this catalytic lysine.

Y51 and E85 were chosen for site-directed mutagenesis studies here because they are consistently among the top POOL-predicted residues across the series. The conservative mutation Y51F maintains the steric bulk of the aromatic side chain but loses the ionizable phenol group. The amide side chain in an E85Q variant can maintain the hydrogen-bonding interactions with the N-H groups of K87 and Y88. D111 is also consistently highly ranked but its structural importance is presumed to be too great to separate from a biochemical role in catalysis; it forms a key salt bridge with R19 and a charge-dipole interaction with T84. The relative positions of residues Y51, K83, E85, and D111 in RA95.5–8F are shown in Figure 2.

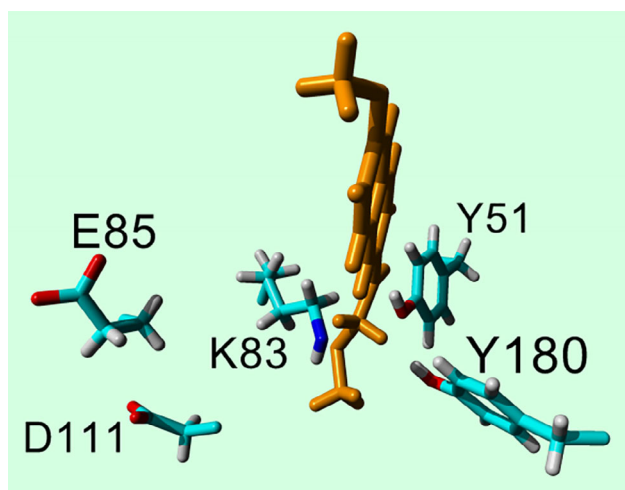


FIGURE 2 Residues Y51, K83, E85, D111, and Y180 in the structure of RA95.5–8F, based on PDB 5AN7.⁹ The ligand is shown in orange. Image rendered in YASARA²⁹

TABLE 2 Turnover^{8,9} rates and μ_4 values for K83 and strongly coupled residues Y51 and Y180 for wild-type variants in the RA95 series; μ_4 values are also shown for the highly ranked acidic residues E85 and D111

Variant	k_{cat} (s ⁻¹) ^a	K83	Y51	Y180	E85	D111
RA95.5	0.00286 ± 0.00016	33 ^b	48 ^b	–	18	31
RA95.5-5	0.172 ± 0.008	37	55	–	19	36
RA95.5-8	0.383 ± 0.024	41	60	–	20	35
RA95.5-8F	11.9 ± 0.60	104	135	70	32	47

^aThese values were measured for the current study and are in agreement with the previously published k_{cat} values.^{8,9}

^bValues averaged for A and B rotamers of K83.

Increasing values for the fourth central moment μ_4 [see Equation (2) above] for the active site residues are observed as the designed enzymes evolve and their activity increases. The μ_4 value is an indicator of the strength of electrostatic coupling between the subject residue and other residues. These calculated μ_4 values are shown in Table 2 for K83 and for the residues most strongly coupled to it, together with the k_{cat} values for each construct, reported previously by Obexer et al.⁹ and measured for the current study. More couplings between the catalytically active residues and nearby residues are built in as the enzyme evolves and thus the μ_4 values rise. For comparison, Table 2 shows the μ_4 values for the highly ranked residues E85 and D111. These values increase in the step from RA95.5–8 to RA95.5–8F as more acidic residues are coupled to them.

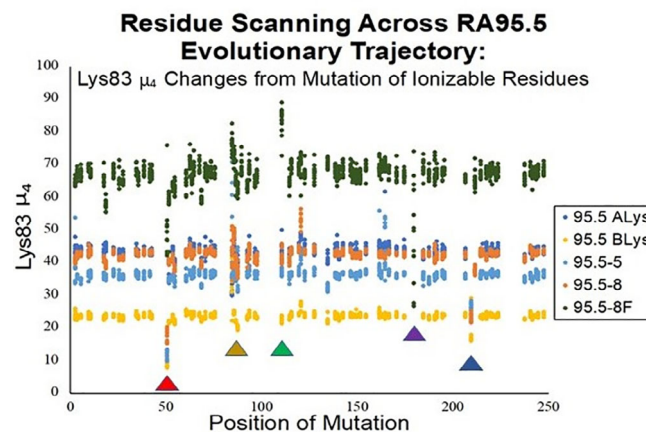


FIGURE 3 Mutational effects on the μ_4 of Lys83 in the evolved variants of RA95. The position of the mutation in the sequence is shown on the x-axis and the calculated μ_4 for Lys83 is on the y-axis. Mutations showing a change from an ionizable residue to a non-ionizable residue are shown. The different colors represent the different structures or variants used: RA95.5 A conformation (dark blue); RA95.5 B conformation (yellow); RA95.5-5 (light blue); RA95.5-8 (orange); and RA95.5-8F (green). Key interacting residues can be identified by the changes they cause the catalytic Lys83. Triangles show sequence positions 51 (red), 85 (yellow), 111 (green), 180 (purple), and 210 (blue)

The combination of residue scanning, where all possible single-site variants are made in silico, and THEMATICS^{12,17,18} shows how different mutations affect the electrostatic features of nearby residues. This provides information to help to identify the new residues or mutations that are responsible for changes in the central moments of K83. Residue scanning results for multiple evolved RA95 family members are displayed in Figure 3. Contributions of specific ionizable residues to the catalytic power of K83 are evident when replacement of a given ionizable residue with any non-ionizable residue type results in a significant change in the calculated μ_4 value of the catalytic K83. Note the substantial reduction in the computed μ_4 value of K83 for all four variants when the tyrosine at position 51 is mutated in silico to a non-ionizable residue. (Sequence position 51 is marked with a red triangle in Figure 3.) The importance of the newly evolved Y180 in RA95.5–8F is also evident, with substantial decreases in μ_4 calculated when this tyrosine is replaced with a non-ionizable residue; sequence position 180 is marked with a purple triangle. Residues R19, Y69, and D165, all conserved through the series, plus a glutamate E135 newly introduced in RA95.5–8F, are all also coupled to K83. The overall trend of increasing μ_4 for Lys83 with increasing activity is also apparent in Figure 3.

4 | EXPERIMENTAL TESTING OF PREDICTED RESIDUES Y51 AND E85

Previously, Y51 was shown to be important in RA95.5–8F⁹ and is predicted to be important in all members of the series studied here. It is believed to function as a proton shuttle in the catalytic mechanism. Y51 and E85 were chosen for experimental investigation here because both scored consistently highly in the POOL rankings throughout the series. To test the effects of the predicted residues Y51 and E85 on the members of the RA95 series, the catalytic activity was measured for variants with mutations of the POOL-predicted residues Y51 and E85. Conservative mutations, that is, Y51F and E85Q, were made. Results of kinetics measurements for the wild type constructs are shown in Table 3, for the Y51F variants in Table 4, and for E85Q variants in Table 5. Results for the

wild type RA95.5–8F and its Y51F variant are consistent with values reported previously.⁹ Michaelis–Menten plots are shown in Supplementary Material.

The loss of Y51 in the Y51F variants results in significant reduction in the μ_4 values for K83 for all members of the series. Table 6 gives the μ_4 values of key residues for homology model structures of the Y51F variants. The E85Q variants have a small effect on the μ_4 values for K83. A graph of $\log_{10}(k_{\text{cat}})$ as a function of μ_4 of K83 for the wild type, Y51F, and E85Q variants is shown in Figure S12 of Supplementary Material.

Thermal shift assays for the wild type and Y51F and E85Q variants are shown in Supplementary Material. In all four constructs, the Y51F variant shows a small reduction in melting temperature of 1–3°C. The effect is larger for the E85Q variants, with reductions in melting temperature of 5–7°C. Melting temperatures are in the range 70–79°C for Y51F and 65–74°C for E85Q.

5 | DISCUSSION

5.1 | Evolution builds stronger coupling between protonation states of active residues

The rising μ_4 values of key residues shown in Table 2 indicate increasing coupling of the protonation equilibria of biochemically active residues with other ionizable residues along the evolutionary trajectory. Strong coupling between protonation states gives rise to expanded buffer ranges,^{19,20,30,31} as in polyprotic acids, enabling the active residues to have significant population of both the protonated and deprotonated states across a wide pH range. Therefore, the side chains can be in the correct protonation state for activity and residues that serve as Brønsted–Lowry acids and bases are able to regenerate for the next turnover cycle.^{19,20}

5.2 | Y51

Table 4 shows that Y51 plays an important role in catalysis for all four constructs, as significant activity is lost

	WT		
	k_{cat} (s ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (M ⁻¹ s ⁻¹)
RA95.5	0.00286 ± 0.00016	279 ± 35	10.3 ± 1.9
RA95.5–5	0.172 ± 0.008	295 ± 32	(5.83 ± 0.90) 10 ²
RA95.5–8	0.383 ± 0.024	275 ± 35	(1.39 ± 0.26) 10 ³
RA95.5–8F	11.9 ± 0.60	308 ± 34	(3.86 ± 0.62) 10 ⁴

TABLE 3 Kinetics constants for wild type members of the RA95 series

TABLE 4 Kinetics constants for Y51F variant members of the RA95 series, with fold change in k_{cat} and in $k_{\text{cat}}/K_{\text{M}}$ compared to wild type

	Y51F				
	$k_{\text{cat}}(\text{s}^{-1})$	$K_{\text{M}}(\mu\text{M})$	$k_{\text{cat}}/K_{\text{M}}(\text{M}^{-1}\text{s}^{-1})$	Fold change k_{cat}	Fold change $k_{\text{cat}}/K_{\text{M}}$
RA95.5	0.00154 ± 0.00028	1700 ± 422	$9.06 \cdot 10^{-7} \pm 3.9 \cdot 10^{-7}$	1.9	11
RA95.5-5	0.00379 ± 0.00017	301 ± 30	$1.26 \cdot 10^{-5} \pm 0.18 \cdot 10^{-5}$	45	46
RA95.5-8	0.00522 ± 0.00011	133 ± 8.5	$3.92 \cdot 10^{-5} \pm 0.19 \cdot 10^{-5}$	73	35
RA95.5-8F	0.162 ± 0.0021	36.7 ± 2.1	$4.41 \cdot 10^{-3} \pm 0.31 \cdot 10^{-3}$	73	8.8

upon mutation to phenylalanine. In RA95.5-5 it is located behind K83 with respect to the substrate, with the OH atom of Y51 situated 4.5 Å away from the NZ atom of K83.⁸ In RA95.5-8F, it is in direct contact with the ligand and the OH-NZ distance is shorter in the crystal structure (3.2 and 4.0 Å for the two rotameric states). Y51 has been identified by molecular dynamics simulation as a residue that contributes to shifting the conformational equilibria toward catalytically competent conformations²⁸; for some states along the dynamic trajectory, partial hydrogen bonding may be possible between K83 and Y51 in the earlier constructs. Even in the absence of direct hydrogen bonding, in the highly evolved RA95.5-8F, K83 and Y51 are connected via a hydrogen bonding network with N110 and Y180. For RA95.5-8F, it is reported to be a part of a catalytic tetrad.⁹ In the proposed mechanism, K83 forms a protonated Schiff base, attacking the carbonyl moiety of the substrate to form a carbinolamine intermediate; Y51 is in position to deprotonate the hydroxyl group of the substrate in the Schiff base intermediate.⁹ In addition, the high, similar intrinsic pK_{a} s of the lysine and tyrosine side chains, with the spatial proximity of Y51 and K83, result in strong coupling of the proton transfer equilibria of the two residues, giving the side chains of both residues a wide buffer range. The wide buffer range enables the primary amine group of K83 to have significant population of its deprotonated state at neutral pH, a feature necessary for attack on the substrate carbonyl group. Calculated energies of interaction between Y51 and K83 are given in Table S1 of Supplementary Material.

The Y51F mutation leads to reductions of 1.9-, 45-, 73-, and 73- fold in k_{cat} compared to wild type for RA95.5, RA95.5-5, RA 95.5-8, and RA95.5-8F respectively, with 11-, 46-, 35-, and 8.8- fold reductions, respectively, in $k_{\text{cat}}/K_{\text{M}}$. The level effect on k_{cat} between RA95.5-8 and RA95.5-8F could be because of a change in rate-determining step along the evolutionary path; the rate-determining step is reported to be early in the mechanism (the first step in Figure 1) for the early constructs, but shifts to the protonation of the enamine intermediate

just prior to product release for the more highly evolved constructs.³²

The introduction of Y180 in proximity to K83 (about 4.4 Å away) in RA95.5-8F, resulting from an F180Y substitution on evolving from RA95.5-8 to RA95.5-8F, adds redundancy to the role of Y51. This could be one contributing factor to the observation that the reduction in k_{cat} is level between RA95.5-8 and RA95.5-8F. Indeed Y180 has been identified as possibly serving as the catalytic base in the highly evolved RA95.5-8F.^{32,33} A previous site-directed mutagenesis study⁹ (Table S1) showed that mutation of either of these two tyrosine residues individually in RA95.5-8F had a relatively modest effect on $k_{\text{cat}}/K_{\text{M}}$, while simultaneous mutation of both tyrosine residues resulted in substantially reduced activity (200-fold loss in catalytic efficiency for the double mutant Y51F/Y180F). Note that Y51, K83, and Y180 are the three top residues in the POOL rankings for RA95.5-8F.

5.3 | E85

E85 is located at the entrance to the active site pocket, on the rim of the beta barrel. Tables 5, 6 show that the conservative mutation E85Q leads to small but significant reduction in catalytic efficiency for RA95.5, RA95.5-5 and RA95.5-8 but has little effect on k_{cat} . It plays some role in thermal stability, as shown in the thermal shift data (Supplementary Material); in RA85.5-8F its side chain hydrogen-bonds to the backbone amide groups of K87 and Y88, helping to stabilize the 85-91 loop. It is one of the acidic residues, D61, E85, D111, E188, located along the base of the beta barrel and conserved across the RA95.5 – RA95.5-8F series. These residues are most likely involved in orienting the substrate in the active site pocket. Additional acidic residues in this location along the rim of the beta barrel appear in the evolutionary trajectory, D212 in RA95.5-8 and E135 in RA95.5-8F. The presence of additional acidic residues adds redundancy and diminishes the role of E85.

TABLE 5 Kinetics constants for E85Q variant members of the RA95 series, with fold change in k_{cat} and in $k_{\text{cat}}/K_{\text{M}}$ compared to wild type

	E85Q				
	$k_{\text{cat}}(\text{s}^{-1})$	$K_{\text{M}}(\mu\text{M})$	$k_{\text{cat}}/K_{\text{M}} (\text{M}^{-1} \text{s}^{-1})$	Fold change k_{cat}	Fold change $k_{\text{cat}}/K_{\text{M}}$
RA95.5	0.00281 ± 0.00021	$1,320 \pm 144$	$2.13 \cdot 10^{-6} \pm 0.39 \cdot 10^{-6}$	1.0	4.8
RA95.5-5	0.0555 ± 0.0034	578 ± 69	$9.60 \cdot 10^{-5} \pm 1.73 \cdot 10^{-5}$	3.1	6.1
RA95.5-8	0.164 ± 0.0071	509 ± 44	$3.22 \cdot 10^{-4} \pm 0.42 \cdot 10^{-4}$	2.3	4.3
RA95.5-8F	10.75 ± 0.50	493 ± 45	$2.18 \cdot 10^{-2} \pm 0.30 \cdot 10^{-2}$	1.1	1.8

TABLE 6 Calculated μ_4 values for key residues in homology models of the Y51F variants

Variant	K83	Y51	Y180	E85	D111	K210
RA95.5 Y51F	17	–	–	22	41	16
RA95.5-5 Y51F	19	–	–	15	24	14
RA95.5-8 Y51F	21	–	–	20	36	19
RA95.5-8F Y51F	36	–	40	33	52	–

6 | CONCLUSIONS

While POOL was created and optimized to predict the biochemically important residues for natural enzymes, it shows predictive capability for this designed enzyme. Note that K83 ranks first in Table 1 for the early construct RA95.5 and ranks first or second for the later constructs. Y51 ranks among the top three for all four constructs. K210 ranks highly for all constructs where it is present. The top three residues for the highly evolved RA95.5-8F are Y51, K83, and Y180, all shown to contribute to catalysis.⁹ N110, previously shown to be important in catalysis,⁹ ranks among the top 12 for RA95.5-5, RA95.5-8 and RA95.5-8F; S110 also ranks eighth in RA95.5.

The calculated values for μ_4 of the catalytic K83, a measure of the degree of coupling between its protonation equilibrium and those of other residues, are 33, 37, 41, and 104, for wild type RA95.5, RA95.5-5, RA95.5-8, and RA95.5-8F, respectively. For the original construct RA95.0, with a measured k_{cat} of $1.0 \cdot 10^{-4} \text{ s}^{-1}$,⁸ its catalytic K210 has a μ_4 of only 15. For contrast, the active site lysine residues of the natural enzyme human fructose 1,6-bisphosphate aldolase³⁴ have μ_4 values of 54, 272, and 190 for K107, K146, and K229, respectively. The larger scaffold of this natural enzyme provides more opportunities for favorable coupling partners to be built in by natural evolution.

Strong coupling between the catalytic K83 and nearby residues, including Y51 and Y180, and also weaker coupling with more distant residues (5–10 Å away), including R19, Y69, and D135, improves the catalytic power of

K83. Thus K83 in the highly evolved enzyme more closely resembles the catalytic lysine residues of natural enzymes than its counterpart in the earlier designs. These couplings expand the buffer range of the side chain and enable the deprotonated state to have significant population at the pH at which the enzyme functions. These couplings are necessary features to build into designed enzymes to produce activity that rivals natural enzymes. Mutations to introduce such couplings can be determined computationally; the difficulty is finding the mutations that do not diminish other vital properties.

7 | MATERIALS AND METHODS

7.1 | Computational

Input structures for POOL predictions were obtained as follows: The RA95.5 (PDB ID 4A2S) and RA95.5-5 (PDB ID 4A2R) structures⁸ were downloaded from the PDB.³⁵ A homology model was built for RA95.5-8, as no structures have been reported. The comparative model was built using the homology model feature in the YASARA³⁶ suite of programs. A total of five models were built using five different templates. The highest-scoring model came from the RA95.5 structure (PDB ID 4A2S).⁸ To evaluate the model, a Ramachandran diagram was drawn using the MolProbity³⁷ server; the model has zero Ramachandran outliers and 98.4% of residues are in favored regions. The model passed the 3D-1D profile assessment by VERIFY3D,³⁸ with 94.4% of the residues at or above the threshold value of 0.2 (>80% is required for the

structure to pass). For RA95.5–8F, the structure (PDB ID 5AN7)⁹ was downloaded from the PDB; the missing loop for residues 58–61 was built in using the homology model module in YASARA. Structures for all four Y51F variants were built using the homology model module in YASARA.³⁶

POOL calculations were performed using electrostatic properties from THEMATICS^{12,18,21} and surface topology properties from the structure-only version of ConCavity³⁹ as input features, using the method of Somarowthu.¹⁵ The average POOL score for all residues in a given structure was used to define the cutoff, similar to the method of Brodtkin⁵ used for natural enzymes; residues with POOL scores higher than the average score are designated, somewhat arbitrarily, as predicted residues for present purposes.

The program FoldEx^{40,41} was used for residue scanning (Figure 3) to generate all possible single-site variants in silico. POOL was performed on the in silico variants as above.

7.2 | Materials

Chemicals, solvents, buffer components, media, and antibiotics were purchased from commercial sources (Sigma-Aldrich, Merck, Acros, Fluka) and used without further purification. Phusion DNA polymerase and the DpnI restriction enzyme were purchased from New England Biolabs (USA). A plasmid purification kit from ZYMO Research (USA) and Ni-NTA agarose from Qiagen (Germany) were used.

7.3 | Site-directed mutagenesis

pET29b plasmids that contain the genes coding for RA95.5, RA95.5–5, RA95.5–8 and RA95.5–8F with a C-terminal His₆-tag were available from previous work.^{8,10,32} The mutations coding for Y51F or E85Q were introduced in all of the templates according to the Quikchange[®] standard protocol (Agilent). DNA oligo synthesis and DNA sequencing were performed by Microsynth AG (Switzerland).

7.4 | Protein production and purification

The 12 different RA95 variants were produced in *E. coli* BL21 Gold (DE3) and purified by nickel affinity chromatography as previously described.^{8,10,32} The enzymes were stored in buffer A (25 mM HEPES pH 7.5 and 150 mM NaCl), which was also used for all kinetic assays.

7.5 | Synthesis of enantiopure (R)-methodol

(*R*)-Methodol was synthesized enzymatically from acetone and 6-methoxy-naphthaldehyde (6-MNA) as reported previously, using RA95.5–8F as the biocatalyst.^{9,32} The reaction mixture containing 2 mM 6-MNA, 2 M acetone and 0.1 μ M RA95.5–8F in 50 ml buffer A was gently shaken for 3 hr at 29°C and subsequently saturated with 15 g solid sodium chloride and extracted with 3 \times 50 ml ethyl acetate. The organic phase was dried over sodium sulfate before removing the solvent under vacuum. The crude material was purified by flash chromatography. (*R*)-Methodol was obtained in 58% yield.¹H-NMR, HR-MS and chiral HPLC were performed for quality control as described previously.^{9,32}

7.6 | Steady-state enzyme activity assays

Kinetic measurements were carried out using a Perkin Elmer UV/Vis spectrometer (Lambda35 with Peltier system and cell changer). (*R*)-Methodol cleavage was measured at 29°C in buffer A supplemented with 2.7% acetonitrile as described previously.^{8,9,32} The formation of 6-MNA was monitored over time by absorption at 350 nm ($\epsilon_{350} = 5,970 \text{ M}^{-1} \text{ cm}^{-1}$). The enzyme concentrations used in the assay ranged from 30 nM for the most active variant to 12 μ M for the least active variant. The (*R*)-methodol concentration in the assay buffer (25 to 900 μ M) was determined by absorption at 330 nm ($\epsilon_{330} = 1,390 \text{ M}^{-1} \text{ cm}^{-1}$) prior to initiating the reaction by addition of enzyme. All measurements were performed in triplicate. The steady-state parameters k_{cat} and K_{M} were determined by fitting the data to the Michaelis-Menten equation.

7.7 | Protein stability measurements using a thermal shift assay

Melting curves were measured according to the standard protocol of the Protein Thermal Shift Kit (Thermo Fisher Scientific) in 96-well PCR plates ($V = 20 \mu\text{L}$ per sample) using a StepOnePlus Real-Time PCR thermocycler (Thermo Fisher Scientific) and SYPRO Orange dye (Thermo Fisher). The melting curves were recorded in triplicate from 25–99°C. The protein concentration was 20 μ M in buffer A. The fluorescence signal derivative was plotted against temperature to extract the midpoints of the melting transitions.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation grants # MCB-1517290 and CHE-1905214, the ETH Zurich, the Swiss National Science Foundation, a National Institute of Justice Predoctoral Fellowship awarded to Timothy A. Coulther, and a Marie Skłodowska-Curie Individual Fellowship (TIMEnzyme) to Cathleen Zeymer. We thank Penny Beuning for valuable discussions.

AUTHOR CONTRIBUTIONS

Timothy A. Coulther: Conceptualization; formal analysis; investigation; writing-review & editing. **Moritz Pott:** Conceptualization; investigation. **Cathleen Zeymer:** Conceptualization; formal analysis; investigation; writing-original draft; writing-review & editing. **Don Hilvert:** Conceptualization; formal analysis; funding acquisition; resources; supervision; writing-review & editing. **Mary Jo Ondrechen:** Conceptualization; formal analysis; funding acquisition; methodology; project administration; resources; software; supervision; visualization; writing-review & editing.

CONFLICT OF INTEREST

The authors declares no conflict of interest.

ORCID

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

REFERENCES

- Kobayashi M, Shimizu S. Metalloenzyme nitrile hydratase: Structure, regulation, and application to biotechnology. *Nat Biotechnol.* 1998;16:733–736.
- 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.
- Mazmanian K, Sargsyan K, Lim C. How the local environment of functional sites regulates protein function. *J Am Chem Soc.* 2020;142:9861–9871.
- Salinas VH, Ranganathan R. Coevolution-based inference of amino acid interactions underlying protein function. *Elife.* 2018;7:e34300.
- Brodtkin HR, DeLateur NA, Somarowthu S, et al. Prediction of distal residue participation in enzyme catalysis. *Protein Sci.* 2015;24:762–778.
- Jiang L, Althoff EA, Clemente FR, et al. De novo computational design of retro-aldol enzymes. *Science.* 2008;319:1387–1391.
- Althoff EA, Wang L, Jiang L, et al. Robust design and optimization of retroaldol enzymes. *Protein Sci.* 2012;21:717–726.
- Giger L, Caner S, Obexer R, et al. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol.* 2013;9:494–498.
- Obexer R, Godina A, Garrabou X, et al. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat Chem.* 2017;9:50–56.
- Obexer R, Pott M, Zeymer C, Griffiths AD, Hilvert D. Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. *Protein Eng Des Select.* 2016;29:355–366.
- Debon A, Pott M, Obexer R, et al. Ultrahigh-throughput screening enables efficient single-round oxidase remodeling. *Nat Catal.* 2019;2:740–747.
- 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.
- Ko J, Murga LF, Wei Y, Ondrechen MJ. Prediction of active sites for protein structures from computed chemical properties. *Bioinformatics.* 2005;21:i258–i265.
- 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 Comp Biol.* 2009;5:e1000266.
- Somarowthu S, Yang H, Hildebrand DG, Ondrechen MJ. High-performance prediction of functional residues in proteins with machine learning and computed input features. *Biopolymers.* 2011;95:390–400.
- Di Cera E, Gill JS, Wyman J. Binding capacity: Cooperativity and buffering in biopolymers. *Proc Natl Acad Sci U S A.* 1988; 85:449–452.
- 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.
- Wei Y, Ko J, Murga LF, Ondrechen MJ. Selective prediction of interaction sites in protein structures with THEMATICs. *BMC Bioinf.* 2007;8:119.
- Shehadi IA, Yang H, Ondrechen MJ. Future directions in protein function prediction. *Mol Biol Rep.* 2002;29:329–335.
- Ringe D, Wei Y, Bojko KR, Ondrechen MJ. Protein structure to function: Insights from computation. *Cell Mol Life Sci.* 2004; 61:387–392.
- Ko J, Andre P, Murga LF, Ondrechen MJ. Statistical metrics for protein active site prediction with THEMATICs. *Proc Fourth Int Conf Bioinfo Genome Regulation Structure BGRS.* 2004;1: 282–285.
- Bajorath J, Kraut J, Li ZQ, Kitson DH, Hagler AT. Theoretical studies on the dihydrofolate reductase mechanism: Electronic polarization of bound substrates. *Proc Natl Acad Sci U S A.* 1991;88:6423–6426.
- Fried, SD, Boxer, SG. Electric Fields and Enzyme Catalysis. *Ann Rev Biochem.* 2017;86:387–415.
- 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.
- Walsh JM, Parasuram R, Rajput PR, Rozners E, Ondrechen MJ, Beuning PJ. Effects of non-catalytic, distal amino acid residues on activity of *E. coli* DinB (DNA polymerase IV). *Environ Mol Mutagen.* 2012;53:766–776.
- Parasuram R, Coulther TA, Hollander JM, Keston-Smith E, Ondrechen MJ, Beuning PJ. Prediction of active site and distal

- residues in *E. coli* DNA polymerase III alpha polymerase activity. *Biochemistry*. 2018;57:1063–1072.
27. Ngu L, Winters JN, Nguyen K, et al. Probing remote residues important for catalysis in *Escherichia coli* ornithine transcarbamoylase. *PLoS ONE*. 2020;15:e0228487.
 28. Romero-Rivera A, Garcia-Borràs M, Osuna S. Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catalysis*. 2017;7:8524–8532.
 29. Krieger E, Vriend G. YASARA view - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics*. 2014;30:2981–2982.
 30. Koumanov A, Rüterjans H, Karshikoff A. Continuum electrostatic analysis of irregular ionization and proton allocation in proteins. *Proteins*. 2002;46:85–96.
 31. Ullmann GM. Relations between protonation constants and titration curves in polyprotic acids: A critical view. *J Phys Chem B*. 2003;107:1263–1271.
 32. Zeymer C, Zschoche R, Hilvert D. Optimization of enzyme mechanism along the evolutionary trajectory of a computationally designed (retro-)aldolase. *J Am Chem Soc*. 2017;139:12541–12549.
 33. Schafer JW, Zoi I, Antoniou D, Schwartz SD. Optimization of the turnover in artificial enzymes via directed evolution results in the coupling of protein dynamics to chemistry. *J Am Chem Soc*. 2019;141:10431–10439.
 34. Dalby A, Dauter Z, Littlechild JA. Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: Mechanistic implications. *Protein Sci*. 1999;8:291–297.
 35. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235–242.
 36. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. *Proteins*. 2002;47:393–402.
 37. Chen VB, Arendall WB III, Headd JJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Cryst D*. 2010;66:12–21.
 38. Eisenberg D, Lüthy R, Bowie JU. *VERIFY3D: Assessment of protein models with three-dimensional profiles*. *Methods in Enzymology*. Amsterdam, Netherlands: Academic Press, 1997; p. 396–404.
 39. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput Biol*. 2009;5:e1000585.
 40. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J Mol Biol*. 2002;320:369–387.
 41. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. *Nucleic Acids Res*. 2005;33:W382–W388.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Coulther TA, Pott M, Zeymer C, Hilvert D, Ondrechen MJ. Analysis of electrostatic coupling throughout the laboratory evolution of a designed retroaldolase. *Protein Science*. 2021;30:1617–1627. <https://doi.org/10.1002/pro.4099>