

# Design and optimization of enzymatic activity in a de novo $\beta$ -barrel scaffold

**Journal Article****Author(s):**

Kipnis, Yakov; Ouald Chaib, Anissa; Vorobieva, Anastassia A.; Cai, Guangyang; Reggiano, Gabriella; Basanta, Benjamin; Kumar, Eshan; Mittl, Peer R.E.; Hilvert, Donald; Baker, David

**Publication date:**

2022-11

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000579643>








**Rights / license:**

[Creative Commons Attribution-NonCommercial 4.0 International](#)

**Originally published in:**

Protein Science 31(11), <https://doi.org/10.1002/pro.4405>

# Design and optimization of enzymatic activity in a de novo $\beta$ -barrel scaffold

Yakov Kipnis<sup>1,2,3</sup>  | Anissa Ouald Chaib<sup>4</sup> | Anastassia A. Vorobieva<sup>1,2,3,5,6</sup>  |  
Guangyang Cai<sup>1,2</sup> | Gabriella Reggiano<sup>1,2</sup>  | Benjamin Basanta<sup>1,2</sup>  |  
Eshan Kumar<sup>1,2</sup> | Peer R.E. Mittl<sup>7</sup>  | Donald Hilvert<sup>4</sup>  | David Baker<sup>1,2,3</sup> 

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, USA

<sup>2</sup>Institute for Protein Design, University of Washington, Seattle, USA

<sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, USA

<sup>4</sup>Laboratory of Organic Chemistry, ETH Zurich, Zurich, Switzerland

<sup>5</sup>VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Brussels, Belgium

<sup>6</sup>Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium

<sup>7</sup>Department of Biochemistry, University of Zurich, Zurich, Switzerland

## Correspondence

David Baker, Institute for Protein Design, University of Washington, Seattle, WA, USA.

Email: [dabaker@uw.edu](mailto:dabaker@uw.edu)

## Funding information

Howard Hughes Medical Institute; Open Philanthropy Project Improving Protein Design Fund; Swiss National Science Foundation

**Reviewing Editor:** John Kuriyan

## Abstract

While native scaffolds offer a large diversity of shapes and topologies for enzyme engineering, their often unpredictable behavior in response to sequence modification makes de novo generated scaffolds an exciting alternative. Here we explore the customization of the backbone and sequence of a de novo designed eight stranded  $\beta$ -barrel protein to create catalysts for a retroaldolase model reaction. We show that active and specific catalysts can be designed in this fold and use directed evolution to further optimize activity and stereoselectivity. Our results support previous suggestions that different folds have different inherent amenability to evolution and this property could account, in part, for the distribution of natural enzymes among different folds.

## KEYWORDS

biocatalysis, computational modeling, enzyme design, enzyme mechanism, protein design

## 1 | INTRODUCTION

Repurposing of natural protein scaffolds for new enzymatic activities is a hallmark of natural evolution and has been used in a multitude of protein engineering

efforts.<sup>1</sup> However, native proteins have complex sequence–structure relationships reflecting their evolutionary history and are often marginally stable.<sup>2</sup> Function-altering amino acid substitutions, which are on average destabilizing,<sup>3,4</sup> can therefore result in unstable protein variants,<sup>5</sup> complicating enzyme design. De novo designed proteins, in contrast, are typically

Yakov Kipnis and Anissa Ouald Chaib equally contributed to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

hyperstable,<sup>6,7</sup> making them more robust to mutation.<sup>3</sup> Because they have more ideal structures, which lack the irregular features (long flexible loops, distorted secondary structure elements, etc.) of most naturally occurring structures, and possess better understood sequence–structure relationships (the principles employed in the design process), they could potentially provide more robust starting points for enzyme design.

The amine-catalyzed retro-aldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (methodol) to 6-methoxy-2-naphthaldehyde and acetone has been widely employed as a model in protein design studies.<sup>8–12</sup> Its mechanism, which involves Schiff base and enamine intermediates, is well understood, the activation barrier that must be overcome is relatively low, and the enzymatic activity can be conveniently detected with spectrophotometric assays. Computational redesign of a range of natural scaffolds has yielded a series of catalysts with useful retro-aldolase activities and selectivities, which have been further improved—sometimes substantially—by directed evolution.<sup>12–17</sup> Optimization often entailed dramatic remodeling of the active site<sup>14</sup> and/or substantial structural changes in the flexible loops flanking the active site,<sup>14,18</sup> neither of which were anticipated in the computational design calculations. De novo scaffolds might provide advantages in this regard over their natural counterparts.

Recent success in the de novo design of hyperstable eight-stranded  $\beta$ -barrels has been exploited to produce receptors for a fluorogenic small molecule,<sup>19</sup> which has led to a new class of fluorescent sensors.<sup>20</sup> While this fold is employed frequently in nature for small molecule binding, it is rarely adopted by natural enzymes. Prostaglandin D synthase,<sup>21</sup> spyrotetronate cyclase AbyU,<sup>22</sup> and allene oxide cyclase<sup>23</sup> are the only known examples. Nevertheless, we reasoned that the high-stability and well-understood sequence–structure relationships of de novo designed  $\beta$ -barrels, together with their ability to host small molecule binding sites, would make them attractive starting points for the creation of new enzymes. Here we report the installation of the requisite catalytic functionality for a retro-aldol reaction into these scaffolds and optimization of the resulting activities by directed evolution.

## 2 | RESULTS

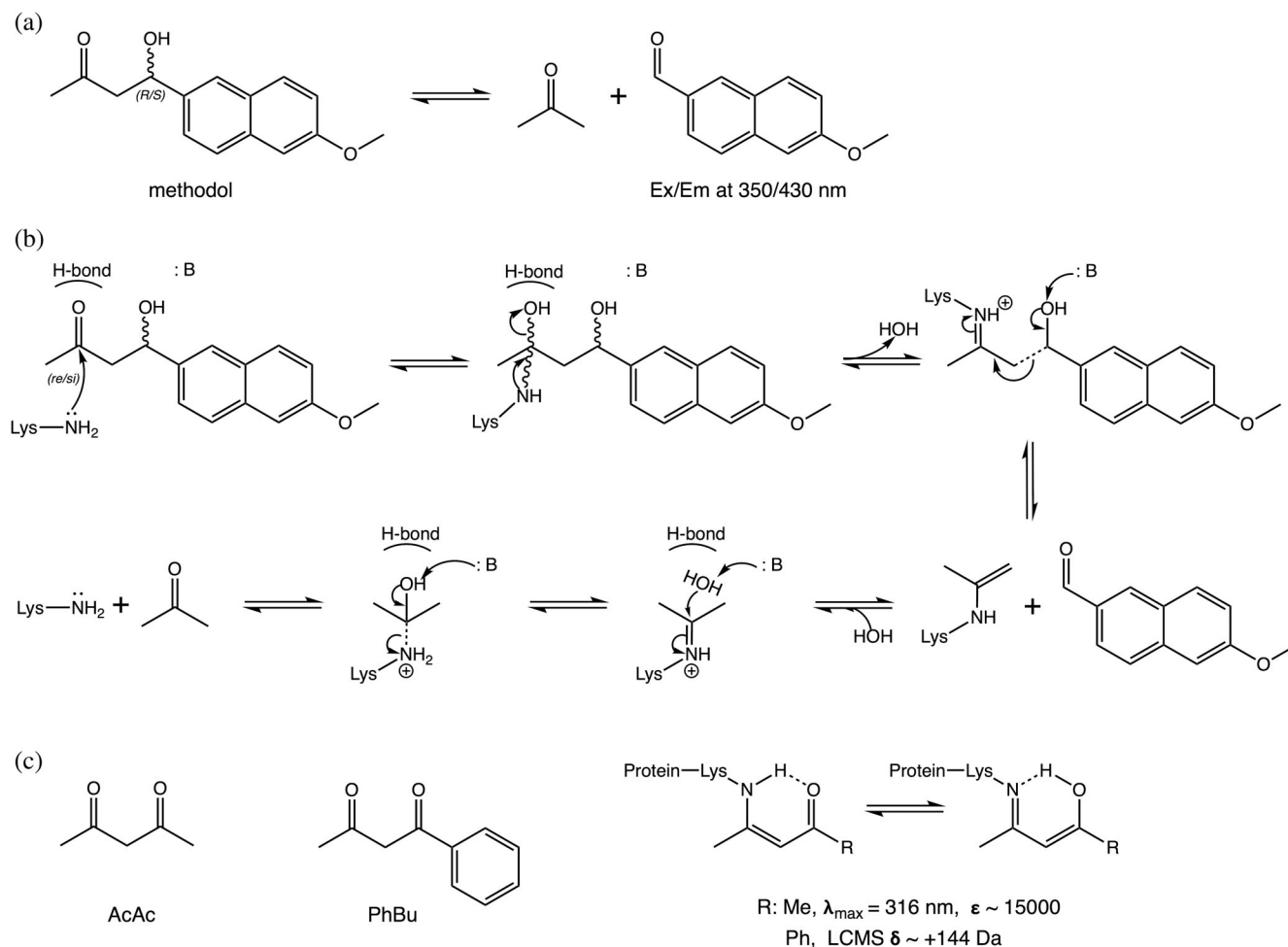
### 2.1 | Protein scaffold design to achieve lysine activation

The amine-catalyzed retro-aldol reaction of methodol is initiated by formation of a Schiff base intermediate

between a lysine residue on the enzyme and the carbonyl group of the aldol substrate (Figure 1a). To function as a nucleophile, the lysine residue must be deprotonated under physiological conditions, which can be achieved by burying its side chain in a hydrophobic environment to lower its  $pK_a$ .<sup>24</sup> We decided to divide the  $\beta$ -barrel scaffold into two regions: the bottom half of the barrel, where the N- and C-termini and the tryptophan corner folding motif<sup>19,25</sup> are located, would serve as the hydrophobic core necessary for proper protein folding and stability, whereas the top half would accommodate the active site (Figure 2a). We reasoned that the Schiff base-forming lysine could be placed in the middle of the  $\beta$ -barrel at the interface between these two regions, projecting from the hydrophobic core into the bottom of an apolar binding pocket.

We generated a series of  $\beta$ -barrel scaffolds de novo based on the blueprint we previously used for the design of small-molecule binders.<sup>19,20</sup> To identify positions in the scaffolds that could accommodate the catalytic lysine, we used a combination of RosettaDesign protocols to systematically replace residues facing the interior of the  $\beta$ -barrel with lysine, place the methodol substrate analog next to the new lysine residue in an orientation consistent with a nucleophilic attack, and redesign the immediate neighborhood of the lysine/substrate complex with hydrophobic residues to create a hydrophobic pocket to increase lysine reactivity and bind the substrate. Five designs with the lowest Rosetta total energy and highest shape complementarity between the substrate and the binding pocket were selected for experimental characterization. For each of these proteins, two different disulfide-stabilized versions were designed and synthesized as codon-optimized genes for expression in *Escherichia coli*. Examples of such designs are shown in Figure S1.

Ten of the designs were produced in *E. coli* and nine could be purified by nickel-affinity chromatography. The only design with a disulfide bond on the water-exposed face of the  $\beta$ -sheet failed to bind to the resin, likely due to aggregation. Eight out of nine designs had circular dichroism (CD) spectra characteristic of  $\beta$ -sheet proteins, and seven of them retained a significant fraction of their secondary structure when heated to 95°C and folded back to the native state upon cooling down to 25°C. Four designs showed a peak on size-exclusion chromatography (SEC) compatible with the retention time of a monomer (Figure S1). To assess the reactivity of the designed lysine, we incubated the purified designs with 1-phenyl-1,3-butanedione ( $M_r = 162 \text{ g mol}^{-1}$ ), which forms a covalent vinylogous amide with activated lysines,<sup>10</sup> and analyzed the products by mass spectrometry. One of the designs (RA $\beta$ b-8, which also appeared to be properly folded based on CD and SEC) showed a peak with the



**FIGURE 1** Retro-aldolase reaction scheme and mechanistic diketone inhibitors used to identify designs with activated nucleophilic lysine residues. (a) Retro-aldolase reaction for methodol proceeds with formation of 6-methoxy-2-naphthaldehyde and can be followed by an increase in fluorescence. (b) Generic reaction scheme of the retro-aldolase reaction catalyzed by the deprotonated lysine side chain amine, complemented by additional polar groups. The reaction proceeds via five covalent intermediates. (c) Diketone inhibitors lead to formation of a stable covalent adduct with the activated lysine. Adduct formation can be measured using spectrophotometry or mass-spectrometry.

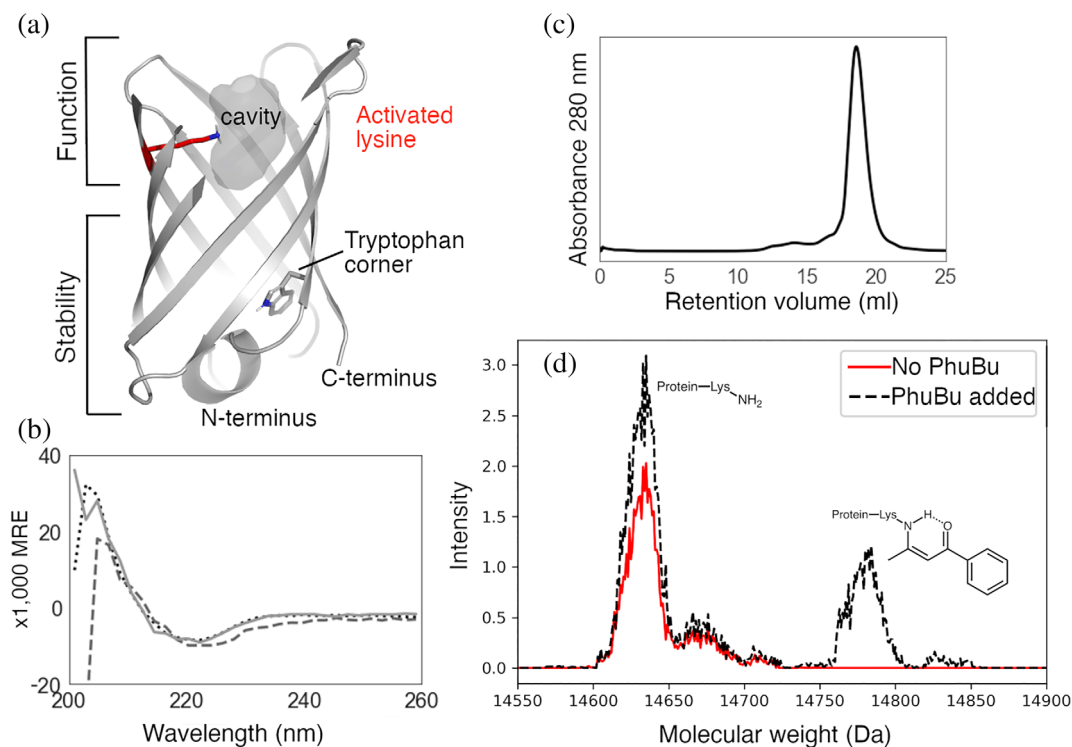
expected 144 Da increase in molecular mass for the adduct with the diketone (Figure 2b–d). The observation of a second peak corresponding to unmodified protein suggests that labeling is only partial, though (Figure 2d).

## 2.2 | Tuning the scaffold to improve lysine reactivity

To improve labeling of the activated lysine in RA $\beta$ -8, we reinvestigated the hydrophobic packing around the lysine and the conformational flexibility of the scaffold. We generated a conformational ensemble using Rosetta Hybridize<sup>26</sup> and found that the hydrogen bonds between the sixth and the seventh strands (accommodating the lysine) partially unzipped in 50% of the lowest energy conformations (Figure 3b,c, left side), allowing escape of the lysine side chain from the hydrophobic environment of the

$\beta$ -barrel interior. Creation of the substrate binding cavity by replacing residues in the hydrophobic core with alanines may have increased the flexibility of the third and fourth hairpins and promoted unzipping of the hydrogen bonds.

To rigidify the interactions between strands 6 and 7, we sought to increase the number of possible backbone hydrogen bonds by elongating hairpin 4 (Figures 3d and S4). To do so, we removed the existing  $\beta$ -turn residues in hairpin 4 and used a loop closure algorithm to identify alternative loops in the PDB with structural compatibility with the edges of the  $\beta$ -strands. We carried out combinatorial sequence design of the grafted loops with consideration of  $\beta$ -turn-specific amino acid preferences. The elongated loop closure solutions with the lowest Rosetta energies fell into two classes. The first comprised classic  $\beta$ -hairpin extensions with equivalent insertions of residue doublets on both sides of the loop. The second class



**FIGURE 2** Biophysical data are consistent with RA $\beta$ -8 being a monomeric, thermally stable and mostly  $\beta$  protein containing a single activated lysine. (a) RA $\beta$ -8 backbone; active site lysine is red. (b) The size-exclusion chromatography profile indicates that GcBar08B is a mostly monomeric protein of approximately 14 kDa. (c) CD spectra of RA $\beta$ -8 measured at 25, 95, and 25°C again after cooling. (d) Liquid chromatography–mass spectrometry spectra of RA $\beta$ -8 incubated with and without the PhBu inhibitor. An extra peak with an approximately +144 Da larger mass indicates formation of the vinylogous amide adduct between the catalytic lysine and diketone, consistent with lysine activation.

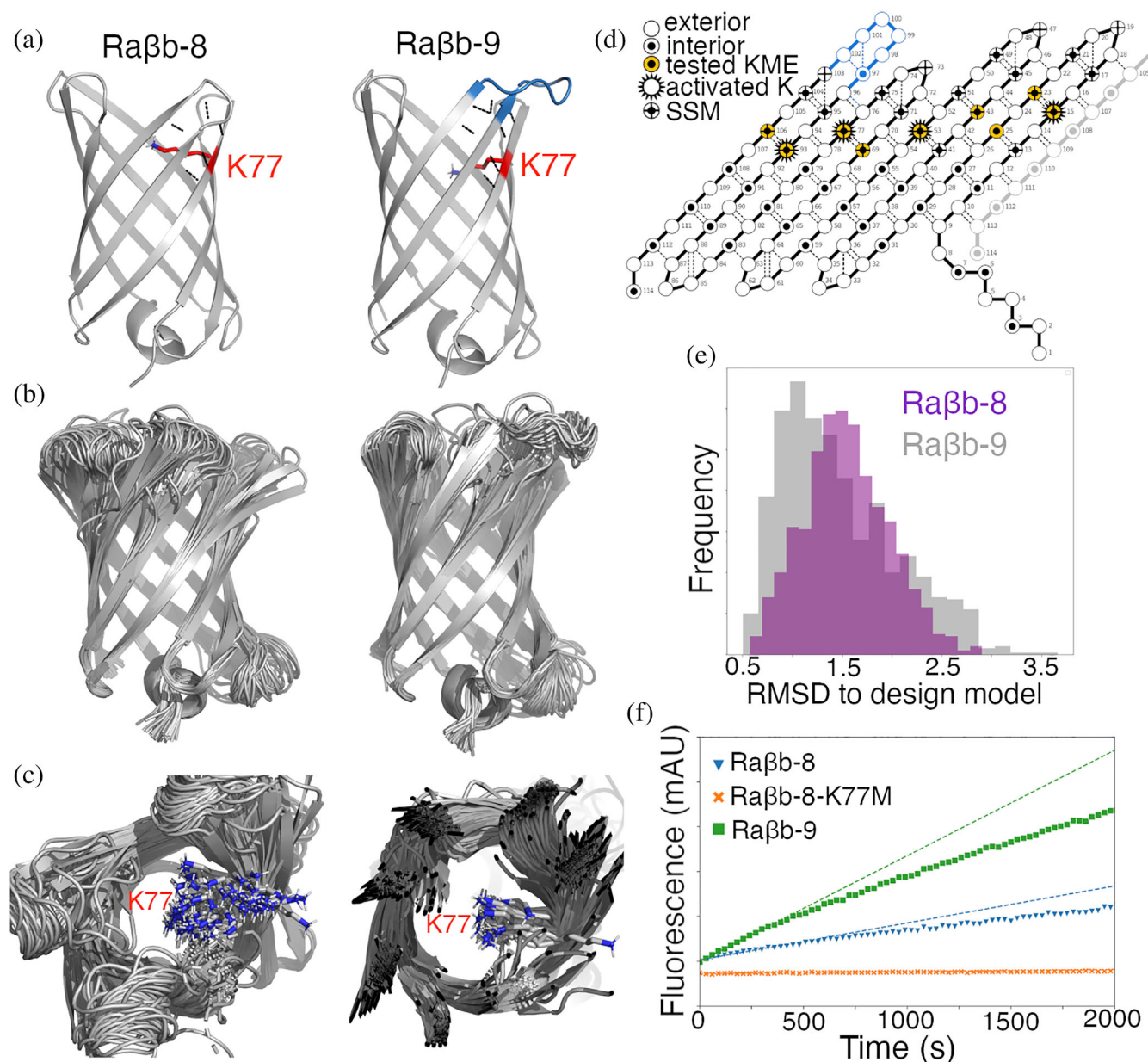
consisted of asymmetric  $\beta$ -hairpin extensions incorporating an additional residue on the eighth  $\beta$ -strand (residue 103 on the 2D map, Figure 3d), forming a new  $\beta$ -bulge. The  $\beta$ -bulge makes an additional hydrogen bond (residue 103 to residue 95, Figure 3d) and has an additional residue pointing toward the core of the  $\beta$ -barrel, which could contribute to higher  $\beta$ -strand rigidity.

Twelve designs with low Rosetta energy and a lysine rotamer that was maintained following perturbation with the Rosetta relax protocol were selected for experimental testing. All 12 designs showed a lower tendency for strand unzipping and lysine escape than RA $\beta$ -8 in conformational ensemble simulations (1–13% of the lowest energy conformations). The designs were expressed, purified, and tested for retro-aldolase activity with racemic methodol. Expression levels of all except two designs were similar or higher than that of RA $\beta$ -8, indicating good tolerance of the scaffold toward loop extension. One of the  $\beta$ -bulge-containing designs, RA $\beta$ -9, which had high structural rigidity (Figure 3b,c, right side and 3e), stood out. In addition to approximately 30% higher expression, it had approximately threefold higher retro-aldolase activity with racemic methodol than the RA $\beta$ -8 design. Although

saturation kinetics were not observed, an apparent  $k_{\text{cat}}/K_M$  value of  $17 \text{ M}^{-1} \text{ min}^{-1}$  was estimated from the initial rate data (Figure 3f), comparable to other computationally designed retro-aldolases prior to optimization by directed evolution (Table 1). These results demonstrate that de novo designed  $\beta$ -barrel scaffolds are amenable to backbone remodeling and can host a reactive lysine residue at the base of a hydrophobic substrate binding pocket.

### 2.3 | Installation of active sites with networks of polar residues

Encouraged by these findings, we set out to design more complete retro-aldolase active sites using RA $\beta$ -9 as a starting scaffold. In addition to a catalytic lysine, the most proficient retro-aldolases generated to date<sup>14,15,17</sup> contain a network of interacting polar residues for transition-state stabilization and proton shuffling. Since many different arrangements of functional groups could potentially perform these tasks, we developed a computational workflow to generate arrays of polar residues around the catalytic lysine to support its attack on methodol. To



**FIGURE 3** (a) Side-by-side comparison of RAβb-8 and RAβb-9. The hairpin 4 extension is in blue, dashed lines are hydrogen bonds. (b) Ensemble of the best scoring structural models computed for RAβb-8 and RAβb-9 suggests larger flexibility of RAβb-8. (c) Top view of the conformational ensemble illustrates catalytic lysine escape due to unzipping of the strands 6 and 7. (d) 2D map of RAβb-9. Extended hairpin 4 is in blue. (e) Histogram plot of RMSD values relative to the design model for the simulations shown in panels b and c. (f) Increased catalytic activity of RAβb-9 compared to RAβb-8. The K77M knockout of RAβb-8 is in orange

ensure that the favorable geometry for this step is preserved throughout the RosettaDesign protocol without resorting to strong geometrical restraints, we modeled the first product of the reaction, a covalent carbinolamine adduct with the catalytic lysine, as a surrogate for the transition state (Figure 4a), and proceeded by sampling conformations of this noncanonical lysine derivative which we denote using the three-letter code KME.

We enumerated rotamer conformations for KME at each position in the upper part of the β-barrel scaffold,

leaving the bottom part of the barrel undisturbed to maintain the overall stability of the protein. Following KME placement in the binding pocket, sets of catalytic groups resembling other retro-aldolase active sites were generated by running the HBNet protocol<sup>27</sup> and requiring KME to be part of the hydrogen-bonded network (see Figure 4c for examples; for a 2D representation of the same designs see Figure S6). While many network configurations were found, inspired by the highly active RA95.5-8F catalyst,<sup>17</sup> we prioritized those that included

TABLE 1 Miscellaneous computationally designed retro-aldolases (assayed with racemic methodol)

Enzyme	$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	$K_{\text{M}}$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{M}^{-1} \text{min}^{-1}$ )	$k_{\text{cat}}/k_{\text{uncat}}$	Ref
Designs					
RA34	0.0073	630	12	$1.9 \times 10^4$	Althoff <sup>13</sup>
RA45	0.0017	800	2	$4.0 \times 10^3$	Althoff <sup>13</sup>
RA60	0.0093	510	18	$2.4 \times 10^4$	Althoff <sup>13</sup>
RA95	0.060	530	11	$1.5 \times 10^4$	Giger <sup>13</sup>
RA110	0.005	1,600	3	$1.2 \times 10^4$	Althoff <sup>13</sup>
RA117	0.0017	473	4	$4.3 \times 10^3$	Bjelic <sup>15</sup>
RA $\beta$ b-9	--	--	17	--	
Evolved					
RA34.6	0.022	30	730	$5.5 \times 10^4$	Althoff <sup>13</sup>
RA45.2-10	0.23	80	2,800	$5.8 \times 10^5$	Althoff <sup>13</sup>
RA60.2	0.070	660	108	$1.8 \times 10^5$	Althoff <sup>13</sup>
RA95.5-8	10.2	200	51,000	$2.6 \times 10^7$	Giger <sup>14</sup>
RA110.4-6	0.29	210	1,400	$5.8 \times 10^5$	Althoff <sup>13</sup>
RA117.4	1.26	373	3,400	$3.2 \times 10^6$	Bjelic <sup>15</sup>

Tyr, Ser or Thr, and Asn or Gln, and filtered out those with Asp or Glu residues within hydrogen bonding distance of the catalytic lysine to prevent formation of a catalytically unproductive salt bridge. We then optimized the residues lining the binding pocket to improve interactions between KME, the polar residues identified by HBNet, and the rest of the scaffold. Because HBNet considers all possible polar interactions compatible with the protein scaffold, this procedure generates a larger diversity of designs than RosettaMatch,<sup>28</sup> which only considers constellations of catalytic groups agreeing with predefined geometrical constraints.

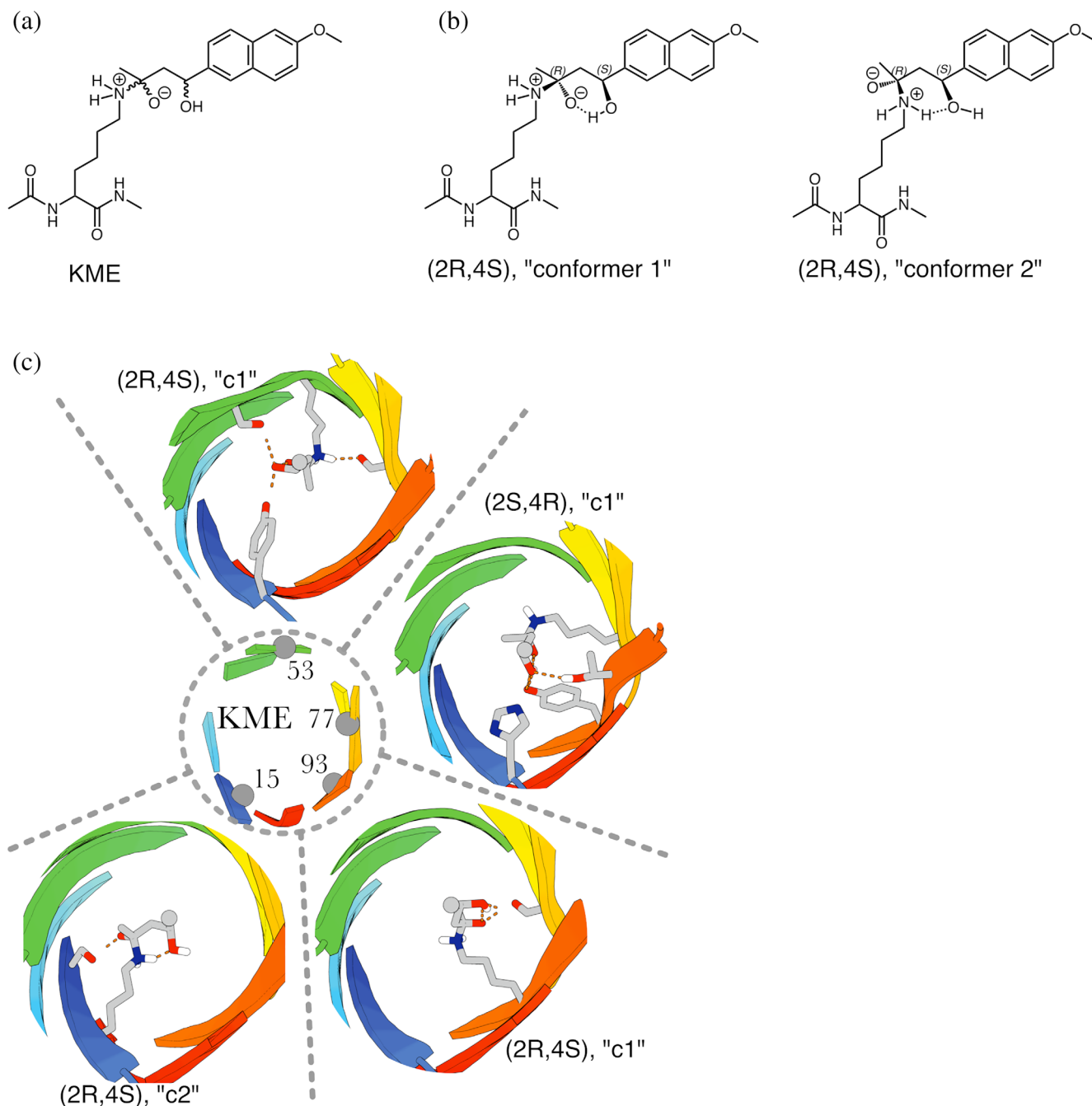
Forty-five of the computational designs (30 using “conformer 1” of the KME and 15 using “conformer 2”, Figure 4b) were expressed, purified, and tested for retroaldolase activity. Six from the “conformer 1” group and one from the “conformer 2” group had detectable retroaldolase activity with racemic methodol. Three of the designs from the “conformer 1” group had the nucleophilic lysine at position 53, two at position 77, and one at position 93. The only active design from the “conformer 2” group had the nucleophilic lysine at position 15. Examples of the designs with measurable activity are shown in Figure 4c. All the designs had lower soluble expression levels than the “parental” RA $\beta$ b-9 protein, probably reflecting decreased stability due to the incorporation of several additional polar residues in the protein core.

Kinetic traces for one variant with a lysine at position 53, RA $\beta$ b-16, exhibited a prominent initial fast phase in

cell lysate assays that quickly leveled off, suggesting high lysine reactivity offset by protein instability under the assay conditions (Figure S3). Indeed, attempts to overproduce this protein for detailed biochemical characterization yielded inclusion bodies. To address this problem, 24 positions in the top part of the  $\beta$ -barrel, including positions designed as part of the putative active site, were individually subjected to saturation mutagenesis (Figure 3d). Mutations improving the apparent enzymatic activity were identified at seven of the targeted sites (A21E, L45D, V49K, A69V, Y75E, V95T, and A104Y). These mutations were combined to afford the variant RA $\beta$ b-16.1, which exhibited greatly improved soluble expression and high activity in crude cell lysates. Four of the substitutions introduce charged residues in the short  $\beta$ -turns connecting consecutive  $\beta$ -strands (Figure 3d), but only two are close enough to the side chains involved in catalysis to affect activity directly. A69V may improve packing of the catalytic lysine at position 53, whereas V95T adds a hydroxyl group that could hydrogen bond to the substrate either directly or via a water molecule.

## 2.4 | Kinetic and structural characterization of RA $\beta$ b-16.1

RA $\beta$ b-16.1 was overproduced, purified, and biochemically characterized. Steady-state kinetic analysis revealed that the enzyme preferentially cleaves (S)-methodol, as expected from the original design model, with a  $k_{\text{cat}}$  of



**FIGURE 4** Unconstrained theozyme generation procedure. (a) Scheme of the lysine–methodol conjugated carbinolamine intermediate used to generate the noncanonical amino acid KME. Backbone capping acetyl and N-methyl groups were used only during initial structure optimization and are not part of the final structure. (b) Two arrangements of the polar atoms in KME, which were considered to promote diversity in hydrogen bonding patterns within the putative active site. A dashed line indicates a hydrogen bond. Only one of four possible stereoisomers for each arrangement is depicted. (c) A number of positions around the circumference of the barrel were individually mutated into one of eight possible variants of KME (two conformers times four stereoisomers) and the HBNet algorithm was applied to design polar residues forming hydrogen bonds with KME and each other. Examples of the generated active sites are shown. The backbone of the protein is colored from blue (N-terminus) to red (C-terminus), naphthyl rings of KME are replaced by a small sphere. The design with (2R,4S)KME (“c1”) at position 53 corresponds to RAβb-16 in the text

1.5 min<sup>-1</sup> and a  $k_{\text{cat}}/K_M$  of 6,500 M<sup>-1</sup>min<sup>-1</sup>. Notably, the selectivity factor  $S$ , given by the ratio  $S = (k_{\text{cat}}/K_M)_S / (k_{\text{cat}}/K_M)_R$ , is 36, which is substantially higher than that

achieved by other computationally designed retroaldolases prior to extensive directed evolution.<sup>12,13</sup> Knockout experiments confirmed that Lys53 is crucial for



TABLE 2 Activity with enantiopure methodol

Enzyme	$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	$K_{\text{M}}$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{M}^{-1} \text{min}^{-1}$ )	$k_{\text{cat}}/k_{\text{uncat}}$	$(k_{\text{cat}}/K_{\text{M}})/k_{\text{Lys}}$	S
S-methodol						
RA $\beta$ b-16.1	$1.5 \pm 0.1$	$230 \pm 40$	6,500	$3.8 \times 10^6$	$4.7 \times 10^7$	36
RA $\beta$ b-16.2	$1.6 \pm 0.1$	$50 \pm 10$	30,000	$4.1 \times 10^6$	$2.2 \times 10^8$	500
Ab 38C2	0.95	22	43,000	$1.8 \times 10^6$	$1.3 \times 10^8$	3,900
R-methodol						
RA95.0	0.003	300	10	$4.8 \times 10^3$	$2.3 \times 10^4$	2.5
RA95.5	0.258	270	955	$6.6 \times 10^5$	$6.9 \times 10^6$	1/3
RA95.5-5	11	410	27,000	$2.8 \times 10^7$	$2.0 \times 10^8$	1/5
RA95.5-8	21.6	230	94,000	$5.5 \times 10^7$	$7.0 \times 10^8$	1/14
RA95.5-8F	648	320	2,000,000	$1.7 \times 10^9$	$1.5 \times 10^{10}$	1/480

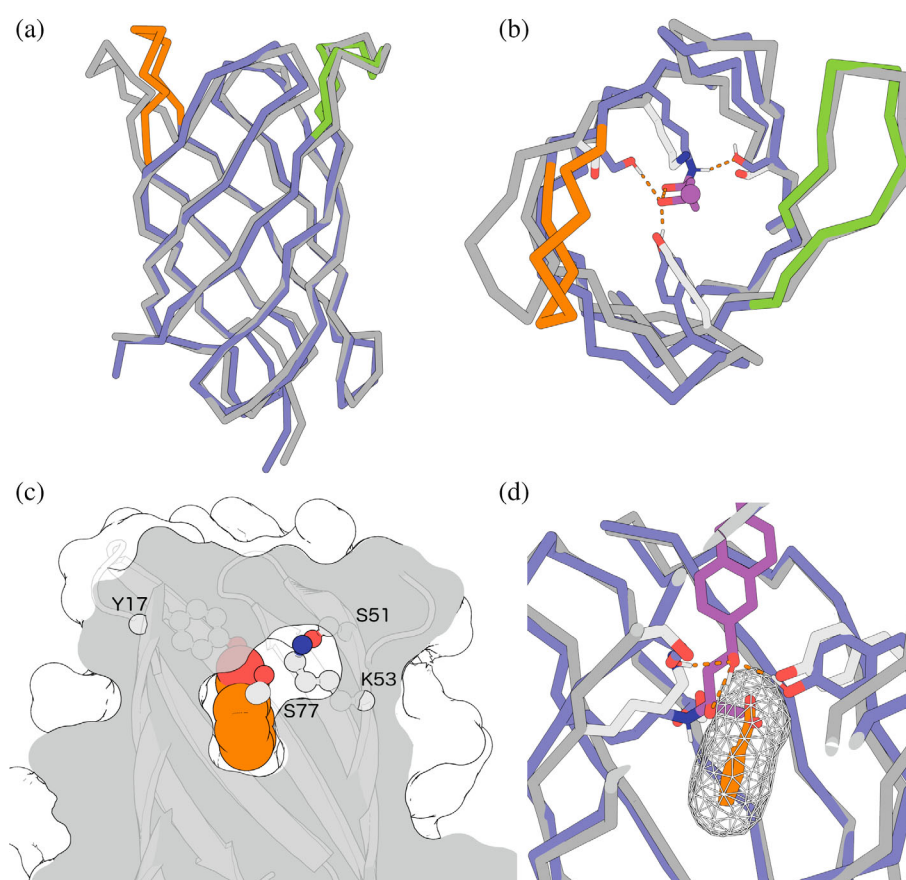


FIGURE 5 Comparison of RA $\beta$ b-16 (model) and the crystal structure of the optimized variant RA $\beta$ b-16.1 (xtal). (a) A backbone trace overlay of the model (purple) and xtal (gray) indicates very good agreement except at the tip of hairpin 2 (orange). The redesigned extension of hairpin 4 (green) shows excellent agreement with the crystal structure. (b) Positions of the designed catalytic side chains are in general agreement with the crystal structure. Atoms corresponding to the methodol part of KME are in magenta (the naphthyl ring of the ligand was replaced by a small sphere for clarity). (c) Benzoic acid (orange) tightly bound in the pocket, active site residues are in spheres (spheres were scaled down for clarity). (d) Benzoic acid binding would interfere with the designed binding mode of methodol, affect conformations of the Lys53 and Tyr17, and disrupt hydrogen bonding.

catalytic activity. Indeed, the steady-state parameters indicate that this residue is  $4.7 \times 10^7$  times more effective as an amine catalyst of methodol cleavage than free lysine in solution (Table 2).

Atomic level insights into the properties of RA $\beta$ b-16.1 were obtained by X-ray crystallography. The structure of the enzyme, determined at 1.8 Å resolution, is very similar to the Rosetta design model with a root mean square deviation (RMSD) of 0.78 Å for all C $\alpha$  atoms (Figure 5a). The elongated loop (hairpin 4) that was designed to

stabilize the protein shows particularly good agreement with the model (RMSD 0.24 Å). The largest difference is localized to hairpin 2, likely because the atypical backbone conformation of Gly53 in the original scaffold was not extensively remodeled when the catalytic lysine was introduced at this site. Consistent with this hypothesis, allowing the original design model to relax by carrying out molecular dynamics simulations improved the agreement with the crystal structure (RMSD 0.44 Å for all C $\alpha$  atoms).

We unexpectedly observed electron density in a small pocket below the catalytic lysine that could be modeled as benzoic acid (Figure 5c). This small molecule, which either copurified with the protein or was introduced as an impurity in the crystallization buffer, forms hydrogen bonds with the HBNet residues Tyr17 and Ser77, as well as with Tyr104. The side chains of Tyr17 and Ser77 adopt the predicted conformations of their counterparts in the design model (side-chain heavy atom RMSD 1.4 Å) and are thus well disposed to react with the substrate and facilitate key proton transfers along the multistep reaction pathway (Figure 5b). Substantial decreases in catalytic efficiency observed upon mutation of Tyr17 provide experimental support for its intended role as a general acid/base.

Although the bound benzoate does not greatly perturb the catalytic apparatus, modeling indicates that it would clash with a carbinolamine intermediate oriented as in the original design (Figure 5d), which placed the long axis of the substrate almost parallel to the long axis of the barrel (Figure S5a). Since control experiments also showed that retro-aldolase activity is not inhibited by benzoate at concentrations up to 2.5 mM (Figure S4), we used docking calculations to explore alternative substrate binding modes. We found that a clash could be avoided by positioning methodol more perpendicular to the long axis of the barrel, placing its methyl group in a pocket adjacent to Val69 and its carbonyl oxygen within hydrogen bonding distance of Thr95, which is backed up by Ser77 (Figure S5b). This pose appears to retain high selectivity for (S)-methodol, which maintains a productive interaction between Tyr17 and the substrate hydroxyl group (Figure S5c), whereas (R)-methodol does not.

## 2.5 | Microfluidic optimization

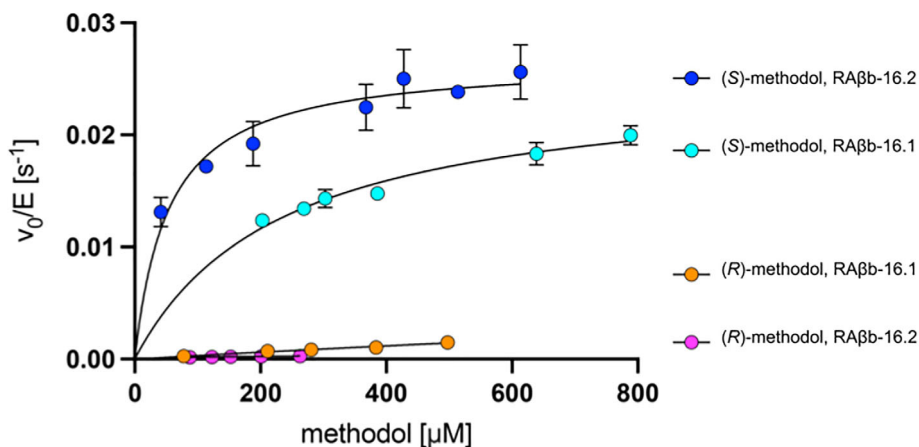
Despite having higher activity than typical first-generation computational designs, RA $\beta$ b-16.1 is still

orders of magnitude less active than RA95.5-8F, the very best artificial retro-aldolase.<sup>17</sup> We therefore turned to laboratory evolution to further optimize its properties. The protein was subjected to error-prone PCR mutagenesis with mutation rates that introduced one to two mutations per gene sequence, and the resulting library was screened for variants that cleave a fluorogenic methodol analogue more efficiently than RA $\beta$ b-16.1 using a microfluidic-based approach based on fluorescence-activated droplet sorting. Previously, this approach was successfully used to evolve RA95.5-8F, which is housed in a natural TIM barrel scaffold, to efficiencies rivaling those of natural class I aldolases.<sup>17</sup>

The most active variant isolated, RA $\beta$ b-16.2, showed approximately threefold higher initial rates than RA $\beta$ b-16.1 in cell lysate assays. The variant RA $\beta$ b-16.2 contains two mutations, K49E and S51H, which are located on  $\beta$ -strand 4 immediately above the catalytic lysine (Figure 5c). Although these mutations did not alter the turnover number ( $k_{\text{cat}} = 1.6 \text{ min}^{-1}$ ), they caused a nearly fivefold drop in  $K_M$  (50  $\mu\text{M}$ ) and a corresponding increase in  $k_{\text{cat}}/K_M$  (30,000  $\text{M}^{-1} \text{ min}^{-1}$ ). Notably, the stereospecificity of the enzyme for (S)-methodol increased in step with catalytic efficiency to give an  $S$  value of 500 (Figure 6). A second round of mutagenesis and screening did not yield further improvements.

## 3 | DISCUSSION

Our work provides the first example of a de novo designed  $\beta$ -barrel enzyme. By creating the protein from scratch using first principles as opposed to redesigning native lipocalins, we avoided any hidden evolutionary bias toward ligand binding that might be associated with native proteins. We further demonstrated that the  $\beta$ -barrel backbone can be computationally remodeled with very high structural precision (Figure 6a,b) to

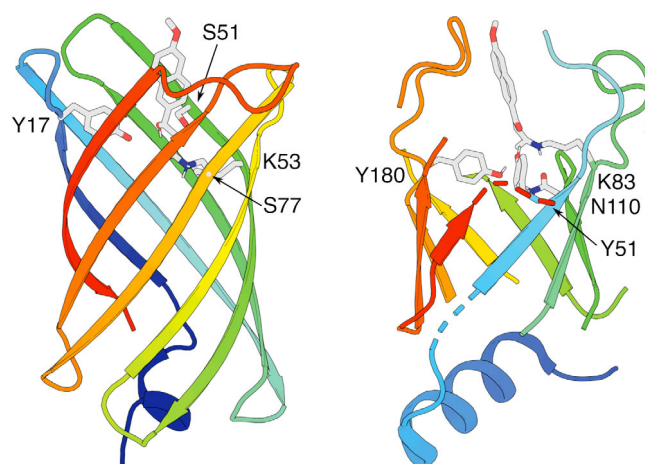


**FIGURE 6** Michaelis-Menten kinetic plots for the cleavage of (S)- and (R)-methodol by RA $\beta$ b-16.1 and RA $\beta$ b-16.2. The reactions were performed at 25°C in 25 mM HEPES, 100 mM NaCl, pH 7 with 5  $\mu\text{M}$  enzyme

customize the scaffold for function, a capability that should be useful for other design efforts as it enables almost unlimited opportunities for backbone exploration and elaboration. Retro-aldolase activity was designed into the  $\beta$ -barrel scaffold by installing a reactive lysine residue in the middle of the barrel to enable amine catalysis, accompanied by a set of polar residues for transition-state stabilization and proton shuffling. The resulting activity, which is comparable to that of the commercially available aldolase antibody 38C2 and as good or better than most computationally designed aldolases in natural protein scaffolds, even after extensive evolutionary optimization,<sup>12,13</sup> attests to the efficacy of this approach. Although the final RA $\beta$ -16.2 variant is not as efficient as the artificial aldolase RA95.5-8F<sup>17</sup> (400-fold lower  $k_{\text{cat}}$  and 66-fold lower  $k_{\text{cat}}/K_M$ ), it exhibits similarly high but complementary stereospecificity, preferentially cleaving (S)- over (R)-configured methodol by a factor of 500. The adventitious binding of a benzoate molecule in a pocket below the catalytic lysine is intriguing considering the natural role of  $\beta$ -barrel proteins as receptors for diverse small molecules. Although the structural data suggest that this ligand may help to preorganize the polar functional groups in the active site, whether it also plays a catalytic role in the retro-aldolase reaction will require further study.

That directed evolution of this  $\beta$ -barrel retro-aldolase leveled off at lower levels of activity than the evolution of RA95.5-8F, which is housed in a TIM barrel scaffold, may hint at possible catalytic limitations of the  $\beta$ -barrel fold. Although diverse enzymatic activities are compatible with TIM barrel folds,<sup>29</sup> including all enzyme-catalyzed aldol reactions in nature,<sup>30</sup> few enzymes utilize  $\beta$ -barrel scaffolds. Why might the  $\beta$ -barrel be a less evolvable scaffold than the TIM barrel? Tawfik and coworkers analyzed structural features contributing to protein fold “innovability,” which they defined as the ability of a fold to acquire a completely new function by divergent evolution,<sup>31,32</sup> and hypothesized that separation of a well-ordered structural core from a conformationally more mobile active site fosters innovation. In this picture, the long, structurally variable loops of the TIM barrel are key to its facile functional diversification, while in our  $\beta$ -barrel based design, the active site and structural core completely overlap, and the rigid antiparallel  $\beta$ -barrel backbone may be less compatible with the dynamics needed for efficient, multistep catalysis (Figure 7).<sup>33</sup>

While the limited functional range of natural  $\beta$ -barrel enzymes supports the idea that such scaffolds may be less “innovable” than TIM barrels, our results show that computational enzyme design can jump-start innovation by equipping these proteins with appropriate functional groups. By moving the active site away from the structural core by increasing the length and the structural



**FIGURE 7** Comparison of retro-aldolase active site locations in RA $\beta$ -16 (left) and RA95.5-8F (right; PDB 5an7; most of the helices and bottom loops of the TIM barrel scaffold are removed for clarity). Protein backbones are colored from blue (N-terminus) to red (C-terminus)

diversity of the loops connecting consecutive  $\beta$ -strands, further design may even be able to bypass inherent limitations on the evolvability of such scaffolds.

## 4 | MATERIALS AND METHODS

### 4.1 | De novo scaffold generation

Sequence-agnostic  $\beta$ -barrel scaffolds were generated using a previously described blueprint<sup>19</sup> and Rosetta BlueprintBDR.<sup>34</sup> To introduce a catalytic lysine into the  $\beta$ -barrel scaffolds, preselected positions in the top part of the barrel were computationally mutated to lysine and proximal sites were subjected to combinatorial sequence design using non-polar residues to improve shape complementarity and hydrophobic packing around the lysine. Disulfide bridges were also incorporated to stabilize the  $\beta$ -barrels containing the newly designed active site cavities using two different approaches. Disulfides between the  $\beta$ -strands of the barrel were designed using the DisulfidizeMover, whereas disulfides between the free N-terminal extension and the main  $\beta$ -barrel domain were designed using the RemodelMover<sup>35</sup> with extensive remodeling of the N-terminus. A combination of computational metrics and visual inspection was used to select the final designs for experimental testing.

### 4.2 | $\beta$ -Hairpin extension

The initial apo-design model was used as input to simulate an ensemble of 1,000 models with resampled

backbone conformations using the HybridizeMover.<sup>26</sup> Each model was relaxed in the torsional and cartesian space. From the lowest-energy models, five models with the lowest  $l_r$ -hb\_bb score ( $\beta$ -sheet backbone-backbone hydrogen bonds) were selected as input for the extension of hairpin 4 together with the original design. In the six input PDBs (the original design plus five alternative conformations simulated with HybridizeMover), the existing hairpin 4 was truncated by five residues (residues 95–99, 2D map on Figure S2). Arg75 was also mutated into alanine in all input models to avoid steric clashes with the extended hairpin. The DirectSegmentLookupMover was used to search a database of protein structures (generated from the PDB) for alternative loops geometrically matching the residues flanking the missing  $\beta$ -turn (residues 93–94 and residues 100–101, 2D map on Figure S4). The rotamers of the grafted loop were allowed to repack, and models were again filtered by long-range hydrogen bond scores. The sequences of the grafted loops were optimized by combinatorial design, biased based on  $\beta$ -turn type (ABEGO type<sup>34</sup>) sequence preferences. Finally, these designs were filtered in two stages: first, for PDBs that had low Rosetta scores; then, for designs in which the catalytic lysine had limited movement during repeated repacking protocols.

### 4.3 | Enzyme design workflow

To enable RosettaDesign protocols for sampling constellations of catalytic residues, Rosetta-style geometry description files were generated for various forms of lysine–methodol carbinolamine adducts. We refer to this noncanonical amino acid by the three-letter code KME. In addition to four possible KME stereoisomers, two side-chain conformations were considered to explore different topologies of the hydrogen bonding network. In “conformer 1” (Figure 4b), the two hydroxyl groups interact via a hydrogen bond, whereas in “conformer 2” (Figure 4b), the 4-hydroxyl group instead interacts with the protonated amino group, as suggested in the work of Houk and coworkers.<sup>36</sup>

Atomic coordinates representing the different KME stereoisomers and torsional conformers were generated with the Avogadro chemical editor, minimized using the built-in Chimera “Minimize Structure” command with parameters assigned by the Antechamber module,<sup>37</sup> and converted into geometry description Rosetta params files using “molfile\_to\_params\_polymer.py” script. The resulting files allow programs in the Rosetta software package to use KME as an additional residue during protein design computations. Rotamer sampling of the non-canonical residue KME during packing was achieved by

generating Rosetta-formatted rotamer libraries following protocol capture in `<Rosetta>/demos/protocol_capture/using_ncaas_protein_peptide_interface_design/HowToMakeRotamerLibraries` as described by Renfrew and coworkers.<sup>38</sup> Alternatively, using a standard lysine rotamer library and specifying the additional rotatable bonds and torsional angle values they sample, allows packability of the KME to be achieved during Rosetta computations.

RosettaScripts xml script was used to perform design protocols where a set of prespecified positions (15, 23, 25, 43, 53, 69, 77, 93, 106) at the top part of the  $\beta$ -barrel were individually mutated to the various KME isomers, HBNet mover was used to find positions in the  $\beta$ -barrel interior that could form hydrogen bonding networks with KME and each other. Finally, FastDesign mover was used to optimize residues interacting with KME and other components of the network to promote packing and desired interactions with the substrate-derived portion of the KME. Designs for experimental testing were selected by a combination of computational metrics, quality of the hydrogen bonding network between tentative active site components, and visual inspection of the structural models.

### 4.4 | Sequence optimization by site-saturation mutagenesis

Individual sites were mutagenized by combining “forward” oligos containing degenerate NDT, VHG, TGG codons at the position of interest in a 12:9:1 ratio with the corresponding “reverse” oligo in an inverse PCR using a circular plasmid with the parental sequence as a template. This strategy utilizes 22 codons to encode 20 amino acids and does not include a stop codon.<sup>39</sup> The resulting linear DNA products with duplicated DNA sequences flanking the ends were directly transformed into competent yeast or bacterial cells. To increase bacterial transformation efficiency by 10–20 fold, the linear DNA product was treated with Gibson master mix to promote circularization of the DNA before transformation was performed.

The following positions were targeted for randomization I13X, A15X, Y17X, N19X, A21X, F23X, I41X, G43X, L45X, N47X, V49X, S51X, K53X, A69X, A71X, T73X, Y75X, S77X, A93X, V95X, V103X, A104X, V106X (Figure 3b). Libraries for each position were transformed separately into XL1Blue or DH5alpha cells for DNA propagation. High quality library DNA was obtained by minipreping 3-ml overnight cultures in lysogeny broth (LB) media with an appropriate antibiotic. Diversification of the desired position was evaluated by the multiplicity

of peaks in Sanger sequencing chromatograms or sequencing of individual colonies obtained by plating a small fraction of the transformation reaction on LB agar plates with an appropriate antibiotic.

#### 4.5 | Selected protein expression and purification

For protein expression, plasmid DNA was transformed into BL21(DE3)Lemo competent cells and a transformation reaction was used to inoculate 3-ml “Terrific Broth” (TB) starter culture complemented by 1% (w/v) glucose and the appropriate antibiotic. An overnight starter culture was diluted 100-fold into 50 ml of autoinduction media complemented with the appropriate antibiotic and the culture was incubated for 24 h at 37°C with shaking. After incubation, cells were harvested by centrifugation for 15 min at 3,000g, pellets resuspended in PBS with 10 mM imidazole and lysed by sonication. Lysate was clarified by centrifugation for 30 min at 20,000g. Supernatant was passed over NiNTA resin in a gravity column, the resin bed was washed with 20 bed volumes of PBS containing 10 mM imidazole, and bound His-tagged proteins were eluted in PBS containing 200 mM imidazole. Eluted proteins were concentrated to approximately 500  $\mu$ l and loaded onto a Superdex75 gel filtration column equilibrated in PBS. Fractions corresponding to monomeric protein were pooled and concentrated to approximately 500  $\mu$ l. Protein concentration was estimated spectrophotometrically by using the absorbance of the protein solution at 280 nm and a molar absorption coefficient approximated from protein sequence using the equation  $\epsilon^{280} = nW \times 5,500 + nY \times 1,490 + nC \times 125$ , where  $nW$ ,  $nY$ , and  $nC$  are the number of Trp, Tyr, and Cys residues, respectively.<sup>40</sup> Retro-aldolase activity was assayed with racemic methodol by mixing equal volumes of protein solution in PBS with a methodol solution in PBS containing 2% acetonitrile. An increase in fluorescence ( $\lambda_{\text{ex}} = 330$  nm and  $\lambda_{\text{em}} = 452$  nm) associated with the formation of 6-methoxy-2-naphthaldehyde was used to monitor reaction progress.

#### 4.6 | Microtiter plate assays

Microtiter plate assays were performed as previously described. Protein production was induced by addition of 2  $\mu$ g ml<sup>-1</sup> tetracycline. Plates were incubated overnight at room temperature and 200 rpm. Cell lysis was carried out with DNase I (NEB) supplemented with the BugBuster Protein Extraction Reagent (Merck) for 1 h at room

temperature with 1,050 rpm shaking speed in 200  $\mu$ l 25 mM HEPES, 100 mM NaCl (Sigma Aldrich), pH 7.5. Upon cell lysis, cell debris was pelleted via centrifugation at 4°C, 4,000 rpm, for 20 min. Retro-aldolase activity was initiated by addition of 50  $\mu$ l of a stock solution of racemic methodol substrate to 150  $\mu$ l supernatant (to give a final concentration of 50  $\mu$ M), and monitored by fluorescence spectroscopy in a Varioscan Microplate reader ( $\lambda_{\text{ex}} = 330$  nm and  $\lambda_{\text{em}} = 452$  nm) for 8 h.

#### 4.7 | Kinetic characterization

Kinetics were determined with methodol and purified protein samples. Measurements were performed at 25°C in assay buffer (25 mM HEPES, 100 mM NaCl, pH 7.5) containing 2.7% acetonitrile. Formation of 6-methoxy-2-naphthaldehyde was monitored spectroscopically at 350 nm ( $\epsilon = 5,970$  M<sup>-1</sup> cm<sup>-1</sup>). Initial velocities were corrected for enzyme concentration and the uncatalyzed background reaction. Steady-state kinetic parameters were obtained by fitting the resulting data to the Michaelis–Menten equation:  $v_0/[E] = k_{\text{cat}}[S]/(K_M + [S])$ , where  $v_0$  corresponds to the initial velocity,  $[E]$  is the enzyme concentration, and  $[S]$  is the substrate concentration.

#### 4.8 | Crystallization and structure determination

RA $\beta$ b-16.1 was crystallized in sitting drops by vapor diffusion with 0.15  $\mu$ l apoprotein (30 mg ml<sup>-1</sup> in 25 mM HEPES, 100 mM NaCl, pH 7.5) and 0.15  $\mu$ l of the 75  $\mu$ l reservoir solution (2 M [NH<sub>4</sub>]2SO<sub>4</sub>, 0.1 M Bis-Tris, pH 6.5) at 4°C for 250 days. The crystal was fished, transferred to a 2  $\mu$ l reservoir drop on a glass slide, and cut in quarters. The crystals were fished from the glass slide and flash-frozen with liquid nitrogen for subsequent x-ray analysis at the Swiss Light Source Synchrotron (Paul-Scherrer-Institute, Villigen, Switzerland). Diffraction data were collected at the PXIII beamline with a PILATUS 2M-F detector using a wavelength of 1 Å in a -173°C nitrogen-gas stream. The data were indexed and integrated with XDS. The structure was solved by molecular replacement with PHASER<sup>41</sup> an ensemble of structures (PDB IDs: 6czg, 6czj, 6d0t). The structure was manually modified in Coot<sup>42</sup> and refined using Phenix.<sup>43</sup> The final refinement with Phenix and corresponding refinement statistics of the RA $\beta$ b-16.1 crystal structure are summarized in Table 3. The structure was deposited in the Protein Data Bank with accession code 8AH9.

**TABLE 3** Crystallographic data and refinement statistics for RAβb-16.1

Wavelength (Å)	1.0
Resolution range (Å)	38.74–1.747 (1.81–1.75)
Space group	C 2 2 21
Unit cell	55.482 77.485 60.652 90 90 90
Total reflections	27,153 (2637)
Unique reflections	13,586 (1326)
Multiplicity	2.0 (2.0)
Completeness (%)	99.96 (99.33)
Mean I/sigma(I)	17.82 (1.28)
Wilson B-factor	24.72
R-merge	0.024 (0.5651)
R-meas	0.03394 (0.7991)
R-pim	0.024 (0.5651)
CC1/2	0.999 (0.565)
CC*	1 (0.85)
Reflections used in refinement	13536(1326)
Reflections used for R-free	675(69)
R-work	0.1865(0.3220)
R-free	0.2257(0.3461)
CC(work)	0.935 (0.679)
CC(free)	0.884 (0.637)
Number of nonhydrogen atoms	1060
Macromolecules	963
Ligands	27
Solvent	70
Protein residues	115
RMS(bonds)	0.010
RMS(angles)	1.64
Ramachandran favored (%)	99.12
Ramachandran allowed (%)	0.88
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.97
Clashscore	9.11
Average B-factor	32.74
Macromolecules	31.44
Ligands	43.83
Solvent	46.30

#### 4.9 | Exploratory docking of methodol in the RAβb-16.1 binding pocket

To evaluate the feasibility of alternative methodol binding modes in the presence of the benzoate molecule, we

used the RosettaScripts xml protocol with the GALigand-Dock mover.<sup>44</sup> The crystal structure model was stripped of all nonprotein atoms except benzoate and relaxed with coordinate restraints applied to backbone atoms to remove Rosetta scoring function artifacts, while preventing excessive changes to the backbone. The (R)- and (S)-methodol enantiomers were added to the pose in a random orientation with the centroid of methodol overlapping with the cartesian coordinates of the ε-amine of Lys53. Distance restraints between the ε-amine of Lys53 and the carbonyl group of methodol were used to keep the substrate in the active site during the docking simulation. An ensemble of low-scoring poses was visually inspected and experimental information about the functional importance of Tyr17, Ser77, and Thr95 and the dispensability of Ser51 was used to distinguish productive binding modes from unproductive ones.

#### AUTHOR CONTRIBUTIONS

**Yakov Kipnis:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (equal); writing – original draft (equal); writing – review and editing (equal). **Anissa Ouald Chaib:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal). **Anastassia Vorobieva:** Conceptualization (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); visualization (equal); writing – review and editing (supporting). **Guangyang Cai:** Investigation (supporting); methodology (supporting). **Gabriella Reggiano:** Formal analysis (supporting); investigation (supporting); methodology (supporting); software (supporting). **Benjamin Basanta:** Investigation (supporting); methodology (supporting); software (supporting). **Eshan Kumar:** Investigation (supporting); methodology (supporting). **Peer R. E. Mittl:** formal analysis (supporting); validation (supporting). **Donald Hilvert:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); resources (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **David Baker:** Conceptualization (equal); funding acquisition (equal); project administration (equal); resources (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

#### ACKNOWLEDGMENT

We thank Stephen Rettie and Xinting Li at University of Washington for help with mass spectrometry experiments.

#### FUNDING INFORMATION

This work was supported with funds provided by Swiss National Science Foundation and ETH Zurich (Anissa

Ouald Chaib, Peer R. E. Mittl and Donald Hilvert), the Howard Hughes Medical Institute (Yakov Kipnis, Anastassia A. Vorobieva, and David Baker), and the Open Philanthropy Project Improving Protein Design Fund (Yakov Kipnis, Benjamin Basanta, Anastassia A. Vorobieva, and David Baker).

### CONFLICT OF INTEREST

The authors declare no competing financial interest.

### DATA AVAILABILITY STATEMENT


Data available in article supplementary material.

### ORCID

Yakov Kipnis  <https://orcid.org/0000-0002-3057-4916>

Anastassia A. Vorobieva  <https://orcid.org/0000-0001-5554-1398>

Gabriella Reggiano  <https://orcid.org/0000-0003-2311-2155>

Benjamin Basanta  <https://orcid.org/0000-0003-1118-5269>

Peer R.E. Mittl  <https://orcid.org/0000-0002-3348-3147>

Donald Hilvert  <https://orcid.org/0000-0002-3941-621X>

David Baker  <https://orcid.org/0000-0001-7896-6217>

### REFERENCES

- Lovelock SL, Crawshaw R, Basler S, et al. The road to fully programmable protein catalysis. *Nature*. 2022;606:49–58.
- Taverna DM, Goldstein RA. Why are proteins marginally stable? *Proteins*. 2002;46:105–109.
- Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The stability effects of protein mutations appear to be universally distributed. *J Mol Biol*. 2007;369:1318–1332.
- Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*. 2009;19:596–604.
- Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. *PLoS Comput Biol*. 2008;4:e1000002.
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003;302:1364–1368.
- Huang P-S, Oberdorfer G, Xu C, et al. High thermodynamic stability of parametrically designed helical bundles. *Science*. 2014;346:481–485.
- Müller MM, Windsor MA, Pomerantz WC, Gellman SH, Hilvert D. A rationally designed aldolase foldamer. *Angew Chem Int Ed Engl*. 2009;48:922–925.
- Tanaka F, Fuller R, Barbas CF 3rd. Development of small designer aldolase enzymes: Catalytic activity, folding, and substrate specificity. *Biochemistry*. 2005;44:7583–7592.
- Wagner J, Lerner RA, Barbas CF 3rd. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science*. 1995;270:1797–1800.
- Mukherjee S, Yang JW, Hoffmann S, List B. Asymmetric enamine catalysis. *Chem Rev*. 2007;107:5471–5569.
- 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.
- Bjelic S, Kipnis Y, Wang L, et al. Exploration of alternate catalytic mechanisms and optimization strategies for retroaldolase design. *J Mol Biol*. 2014;426:256–271.
- Obexer R, Studer S, Giger L, et al. Active site plasticity of a computationally designed retro-aldolase enzyme. *ChemCatChem*. 2014;6:1043–1050.
- 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.
- Nestl BM, Hauer B. Engineering of flexible loops in enzymes. *ACS Catal*. 2014;4:3201–3211.
- Dou J, Vorobieva AA, Sheffler W, et al. De novo design of a fluorescence-activating  $\beta$ -barrel. *Nature*. 2018;561:485–491.
- Klima JC, Doyle LA, Lee JD, et al. Incorporation of sensing modalities into de novo designed fluorescence-activating proteins. *Nat Commun*. 2021;12:856.
- Yoshihiro U, Toshiki T, Naomi E, et al. Structural and functional significance of cysteine residues of glutathione-independent prostaglandin D synthase: Identification of CYS65 as an essential thiol. *J Biol Chem*. 1995;270:1422–1428.
- Byrne MJ, Lees NR, Han L-C, et al. The catalytic mechanism of a natural Diels-Alderase revealed in molecular detail. *J Am Chem Soc*. 2016;138:6095–6098.
- Neumann P, Brodhun F, Sauer K, et al. Crystal structures of *Physcomitrella patens* AOC1 and AOC2: Insights into the enzyme mechanism and differences in substrate specificity. *Plant Physiol*. 2012;160:1251–1266.
- Isom DG, Castañeda CA, Cannon BR, García-Moreno B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc Natl Acad Sci USA*. 2011;108:5260–5265.
- Greene LH, Hamada D, Eyles SJ, Brew K. Conserved signature proposed for folding in the lipocalin superfamily. *FEBS Lett*. 2003;553:39–44.
- Song Y, DiMaio F, Wang RY-R, et al. High-resolution comparative modeling with RosettaCM. *Structure*. 2013;21:1735–1742.
- Boyken SE, Chen Z, Groves B, et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science*. 2016;352:680–687.
- Zanghellini A, Jiang L, Wollacott AM, et al. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci*. 2006;15:2785–2794.
- Nagano N, Orengo CA, Thornton JM. One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol*. 2002;321:741–765.
- Goldman AD, Beatty JT, Landweber LF. The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J Mol Evol*. 2016;82:17–26.
- Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? *Fold*

- polarity and stability trade-offs. *J Mol Biol.* 2013;425:2609–2621.
32. Tóth-Petróczy A, Tawfik DS. The robustness and innovability of protein folds. *Curr Opin Struct Biol.* 2014;26:131–138.
  33. Richard JP. Protein flexibility and stiffness enable efficient enzymatic catalysis. *J Am Chem Soc.* 2019;141:3320–3331.
  34. Koga N, Tatsumi-Koga R, Liu G, et al. Principles for designing ideal protein structures. *Nature.* 2012;491:222–227.
  35. Huang P-S, Ban Y-EA, Richter F, et al. RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS One.* 2011;6:e24109.
  36. Bahmanyar S, Houk KN. Transition states of amine-catalyzed aldol reactions involving enamine intermediates: Theoretical studies of mechanism, reactivity, and stereoselectivity. *J Am Chem Soc.* 2001;123:11273–11283.
  37. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 2006;25:247–260.
  38. Douglas Renfrew P, Choi EJ, Bonneau R, Kuhlman B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide Interface design. *PLoS One.* 2012;7:e32637.
  39. Kille S, Acevedo-Rocha CG, Parra LP, et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth Biol.* 2013;2:83–92.
  40. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 1995;4:2411–2423.
  41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst.* 2007;40:658–674.
  42. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004;60:2126–2132.
  43. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr D Struct Biol.* 2019;75:861–877.
  44. Park H, Zhou G, Baek M, Baker D, DiMaio F. Force field optimization guided by small molecule crystal lattice data enables consistent sub-angstrom protein-ligand docking. *J Chem Theory Comput.* 2021;17:2000–2010.

## SUPPORTING INFORMATION

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

**How to cite this article:** Kipnis Y, Chaib AO, Vorobieva AA, Cai G, Reggiano G, Basanta B, et al. Design and optimization of enzymatic activity in a de novo  $\beta$ -barrel scaffold. *Protein Science.* 2022; 31(11):e4405. <https://doi.org/10.1002/pro.4405>