Computational design of a thermostable de novo biocatalyst for whole cell biotransformations

Wael Elaily[1], David Stoll[1], Morakot Chakatok[1], Matteo Aleotti[2], Birgit Grill[1], Horst Lechner[1,3], Mélanie Hall[2,3,4] and Gustav Oberdorfer*[1,3]

[1]Institute of Biochemistry, Graz University of Technology, Petersgasse 12/2, 8010-Graz, Austria

[2]Institute of Chemistry, University of Graz, Heinrichstrasse 28, 8010 Graz Department of Chemistry, University of Graz, Austria.

[3]BioTechMed Graz, Mozartgasse 12/2, 8010-Graz, Austria

[4]BioHealth, Field of Excellence, University of Graz, Austria

*Correspondence to: gustav.oberdorfer@tugraz.at

Abstract

Several industrially relevant catalytic strategies have emerged over the last couple of decades, with biocatalysis gaining lots of attention in this respect. However, this type of catalysis would be even more thriving if customizable and stable protein scaffolds would be readily available. Such highly stable protein scaffolds could act as asymmetric reaction chambers in catalyzing a wide range of reactions. In this study, we are detailing the design and experimental characterization of computationally designed de novo proteins with a non-natural fold, which act as biocatalysts. The initial design and several variants of it, catalyze the aldol condensation and retro-aldol reaction. All designs form a helical barrel structure comprised of six antiparallel straight helices connected by five loops, resulting in an open channel with two accessible cavities. The designs exhibit high thermal stability and an excellent overall fit between measured and calculated scattering profiles from small-angle x-ray scattering experiments. An experimentally determined crystal structure of a surface redesigned variant shows close to perfect agreement to the design model. To highlight the versatility and mutational tolerance of this fold, we rationally designed variants in which the active site is placed at different positions throughout the central channel of the original design. Finally, we showcase that the initial design as well as variants exhibit significant aldolase/retro-aldolase activity in whole cell biotransformations, making them the first de novo biocatalysts to be tested in this form. With a tolerance of up to 20% of organic solvent, these designs hold promise for further utilization and optimization in the field of white biotechnology.

Introduction

Since the turn of the century, biocatalysis has emerged as an indispensable tool for chemists, offering unparalleled precision and selectivity in chemical transformations[1]. Particularly in the production of fine chemicals, biocatalysis has become a method of choice, owing to its ability to finely control stereochemical outcomes of reactions while producing minimal amounts of waste and exhibiting an excellent resource-balance[2]. Early efforts in biocatalysis predominantly relied on natural enzymes, which often posed limitations in terms of scope and versatility. However, with the advent of directed evolution, a revolutionary opportunity arose to tailor native enzymes to specific process conditions. This breakthrough facilitated the enhancement of enzyme activity for both native and novel substrates and the optimization of stereoselectivity and adaptability to diverse reaction environments[3,4]. In addition to these experimental advancements, computational tools offered valuable assistance in the engineering and analysis of experimental-directed evolution workflows The integration of computational methods into the biocatalysis toolkit represented a significant milestone, promising to expedite the development of next-generation biocatalysts with enhanced performance characteristics[5]. Moreover, the use of biocatalysts for synthetic purposes has contributed significantly to the transition of the chemical and pharmaceutical industries towards clean production processes and a renewable resource base[6]. Most of these approaches relied either on the identification of a spurious activity in a naturally occurring protein or used natural protein scaffolds as starting points. In both situations the starting points likely are compromised by evolutionary effects[7], something that can be overcome by using de novo proteins[8–11].

Custom designed de novo enzymes can become an alternative to natural starting points and should even further enhance the aforementioned advances to aid in the remedy of biotechnological challenges of the 21st century. Several recent publications demonstrated this, creating highly active luciferases, heme-dependent (oxidation) enzymes, xylanases, p450 enzymes and more[12–16]. An emphasized detail of these studies is their high specificity for the designed reaction and substrates. As a result, though, the current paradigm to design efficient biocatalysts is to design entirely new proteins for each new substrate and reaction. While this approach is valuable and robust, it is not suited to quickly engineer catalysts for well-known chemical reactions or reactions accessible only through cofactor binding. Additionally, small de novo scaffolds often are not very mutation tolerant[17,18]. For this, we need to address functional design and selection for activity within a single, highly mutational load resistant protein fold. The broad availability of accurate structure prediction methods and protein design tools encourages an investigation in this direction. Research in this direction also offers a shot to apply de novo

biocatalysts for whole cell biotransformations, arguably still the most valuable way of using enzymes for synthesis purposes.

Here, we present a highly versatile protein scaffold which acts as an efficient whole cell biocatalyst, designed by finely sampling helical arrangements of long straight alpha-helices using parametric design[19]. In addition to finding optimal packing solutions for an anti-parallel six helix barrel with high thermostability, we investigate placement of a potent catalytic array along the central channel of the design to probe its resistance to high mutational loads and exemplify its versatility as a re-usable asymmetric reaction chamber by designing nine variants. Biochemical and biophysical characterization as well as an experimentally determined structure allowed us to test several of the above-described features, we deem necessary for a universally applicable de novo protein backbone for biotransformations.

Results

Parametric design allows for ultrafine backbone geometry optimization

We set out to design a multifunctional de novo protein chassis, which can be used as an asymmetric reaction chamber. To do so, we aimed at designing anti-parallel six-helical barrels accommodating ligand or substrate accessible channels, starting from the design principles elucidated during the design of four-helix bundles[19]. To identify stable arrangements, we extensively sampled super-helical coiling, radius and the translation along the central axis of the individual helices. Loops to connect the individual helices were kept short and idealized. This resulted in lowest energy sequences for totally straight helical arrangements with 18-residues repeating units and an average diameter of 15.7 Å ($C\alpha$-$C\alpha$ of opposing helices, Figure 1A). In contrast to smaller helical bundles, the large helical barrels do not allow for tightly packed cores. Thus, the biggest contribution to structural integrity is coming from neighboring anti-parallel helices. Naturally occurring helical barrels with similar numbers of anti-parallel helices typically feature oligomerization of smaller, stable entities or stabilizing domains on each side of the helix barrel due to the lack of a stably packed core[20], which makes this approach very challenging from a design perspective. The backbones of these designs were generated as described previously[21] and by using the parametric design code as implemented into Rosetta[22]. The straight helices and overall configurations made it possible to install salt-bridges across the barrel at two distinct positions, thereby, enhancing the folding and stability of the generated designs (Figure 1A).

We generated eleven versions of such designs with different twists and repeating units, however, only one of the designs – 6H5L – could be produced in high amounts in the wet lab and exhibited an all-helical circular dichroism (CD) signal and high thermostability (Figure 1C&D).

Interestingly, during combinatorial sequence design calculations, we found that the frequency of incorporation of alanine at helical interface positions strongly depended on the set supercoil and individual helical coiling, with some designs being overpopulated with this residue type, even though no obvious geometry or space restrictions were apparent. Therefore, we adjusted the Rosetta reference energies[23] to achieve the desired residue identity balance. All the herein described designs were computed prior to the advent of machine learning based structure prediction methods and thus could not be tested in this respect prior to ordering synthetic genes. However, we and others recently found that AlphaFold2[24,25] could accurately predict the structures of designed helical barrels from single-sequence input[26]. Therefore, we checked 6H5L and all its variant sequences subsequently on their prediction quality (suppl Figure 4).

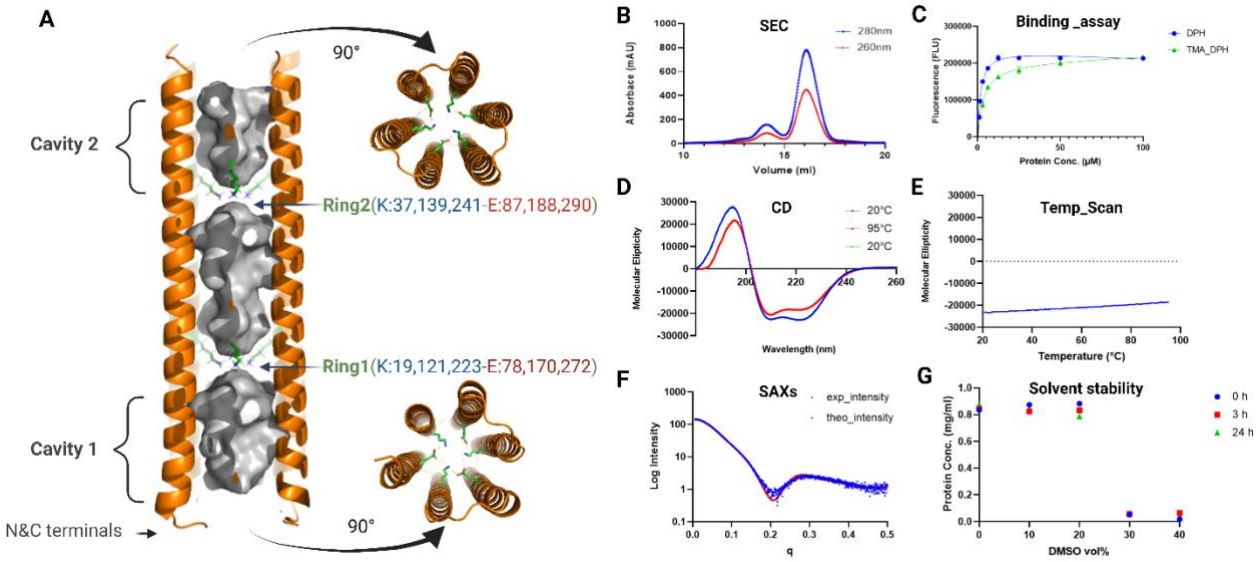

Figure 1. Characterization of 6H5L design **A)** Structure representation consists of a six-helix barrel, with two rings introduced into the hydrophobic channel, creating two cavities, each consists of three lysine (K) and three glutamate (E) residues connected by electrostatic interactions, contributing to structural stability and enhanced catalytic efficiency. **B)** Size exclusion chromatogram (SEC) of 6H5L desired monomer fraction as the highest peak at the expected retention volume with good separation from the dimer form. **C)** Saturation binding curves of different 6H5L protein concentrations with 1μM constant concentration of DPH or ATM-DPH fluorescent dyes. This result indicates that the protein folds in a barrel shape giving rise to a hydrophobic channel. **D)** Circular dichroism (CD) spectrum confirmed an all alpha-helical fold. **E)** CD Temperature scan exhibited high thermal stability with only small changes until 95°C with ideal refolding by cooling back to 20°C. **F)** SAXS solution scattering plot showing good fitting between the theoretical (red) and measured (blue dots) curves with chi-square of 2.2. Curve fitting was executed using the FoXS online server. **G**) Solvent stability was measured by determining the protein concentration using absorbance at 280 nm, plotted against DMSO percentage at three time points (0h in red, 3h in light blue, and 24h in gray), indicating protein stability over time up to 20% DMSO.

Characterization of the parametrically designed helix bundles in *E. coli*.

Production of the parametrically designed protein 6H5L and all its variants yielded soluble protein after production in *E. coli.* The proteins were purified using affinity and size-exclusion chromatography, yielding nearly identical SEC chromatogram profiles for 6H5L and its variants, featuring a prominent absorbance peak at 280 nm with the expected retention volume for a protein of this size. In addition, the chromatograms displayed a second minor peak, which corresponded to a dimeric form of our designs. However, excellent separation between the monomeric and dimeric form could be achieved (Suppl Figure S2A). The pooled and concentrated fractions from the monomer peak were confirmed to contain highly pure protein (Suppl Figure S2B).

To confirm folding and the secondary structure content of our designs, we recorded CD spectra for both pure 6H5L (Figure 1D) and its variants (Suppl Table S6). All recorded spectra exhibited a characteristic α-helical profile. Notably, these structures exhibit very high thermal stability as exemplified by only a minimal loss of signal (222nm) during temperature ramps up to 95 °C (Figure 1C, 1D) and refolding after cooling back down to 20°C. This highlights the robustness of the designs, even when introducing mutations in the core for active site redesign (all single K and YQY variants) or on the surface for improved crystal formation (FKQ variant). A summary of all CD spectra from all variants can be found in Supplementary Table S6. To assay barrel formation, we resorted to previously described binding experiments of fluorescent molecules.[27] One of the benefits of this assay is that it can be performed in a semi-high throughput manner using 96-well plates. All designs exhibited saturation binding curves, which could be fitted to a one-site total binding model for calculating dissociation constants (K_D). The observed affinities are in the micromolar range with specific K_Ds between 1.1 to 4.5 μM (Figure 2B and Supplementary Table S6). These results corroborate the designed barrel-like structures for all designs. To further investigate the correct folding and dynamics of our designed proteins, we employed SAXS measurements. The obtained results show good fits between the measured and calculated scattering profiles, characterized by low chi² values, particularly within the Guinier region of the logarithmic plot (Figure 2F and Suppl Table S6). All obtained fits indicate that the designs exhibit the expected behavior in solution, consistent with the AlphaFold2 predicted models[25,28].

An important feature of a practically applicable biocatalyst is its stability and activity in organic (co)solvent. We assessed the stability of 6H5L *in* up to 30% DMSO over time points of 0, 3, and 24 hours. Notable instability was observed only in concentrations of 30% DMSO and higher, highlighting the proteins high solvent tolerance (Figure 2G).

Protein crystallization and structure determination

Initial crystallization of the original 6H5L sequence resulted in the formation of 2D crystals only and ultimately did not produce high enough diffraction quality to determine an experimental structure. However, protein modification to enhance crystal contact formation by exchanging a total of 15 lysine residues on the equivalent helical wheel *f* position with glutamine (6H5L_FKQ variant), lowered the pI from 8.99 to 5.7 and resulted in well-ordered three-dimensional crystals diffracting up to 2.8 Å. To ensure that this extensive residue exchange did not interfere with correct folding, we used AlphaFold2 to predict the 6H5L_FKQ structure and found that it closely matched the original 6H5L, with an RMSD of 0.17 across all atoms.

The 2.8 Å structure of 6H5L_FKQ shows an overall high degree of similarity to that of the design model (Figure 2) with the predicted straight helices, no supercoil twist and the 18-residue repeating unit. The packing differences within individual repeats between design and structure are minimal with an Ca-atom RMSD of ~0.7 Å over the core repeating units. This structure is another example of a computationally designed helix bundle that deviates from perfect supercoil geometry, further corroborating that low energy structures for monomeric antiparallel helix bundles are not strictly confined to the space spanned by the Crick parameterization, particularly not near the turns.

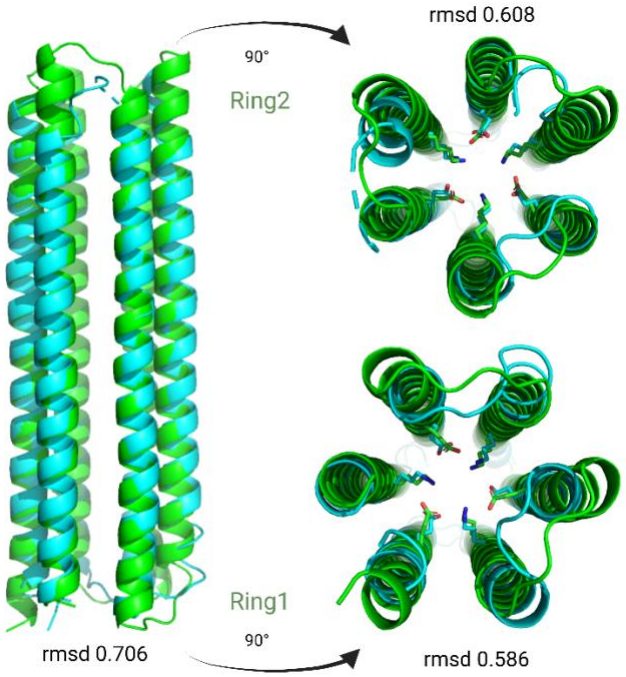

Figure 2. Superposition of parametrically designed 6H5L (in green) and the experimental crystal structure of 6H5L_FKQ (in cyan), highlighting the high overall similarity between model and structure. Residues constituting the two stabilizing electrostatic rings in the center of the structure adopt highly similar geometries in the design and the experimental structure.

Exploring the catalytic potential of the *de novo* designed structure.

The design model of 6H5L, confirmed by crystallography, presents an open, straight six-helix barrel with a hydrophobic channel, highly effective at binding hydrophobic aromatic molecules, as proven by DPH and TMA-DPH fluorescence assays. Crystallographic analysis also identified two hydrophobic cavities within the barrel, each accommodating three lysine residues, electrostatically coupled with three glutamates, forming two identical electrostatic rings. We reasoned that this feature should in principle enhance the lysine nucleophilicity, which is crucial for various enzymatic reactions. Moreover, structure prediction and experimental characterization of variants, with a single lysine in the central channel and the remaining five replaced by arginines, still showed very similar behavior to the original 6H5L design. This enables the identification of the active lysine residue. Have a potentially nucleophilic lysine present, we opted to assay the designs for the retro-aldol reaction, utilizing Methodol **3** as a substrate (Figure 3A). This was selected as the model system due to its simplicity and the established reliability and sensitivity of fluorescence measurement techniques, facilitating easy application.

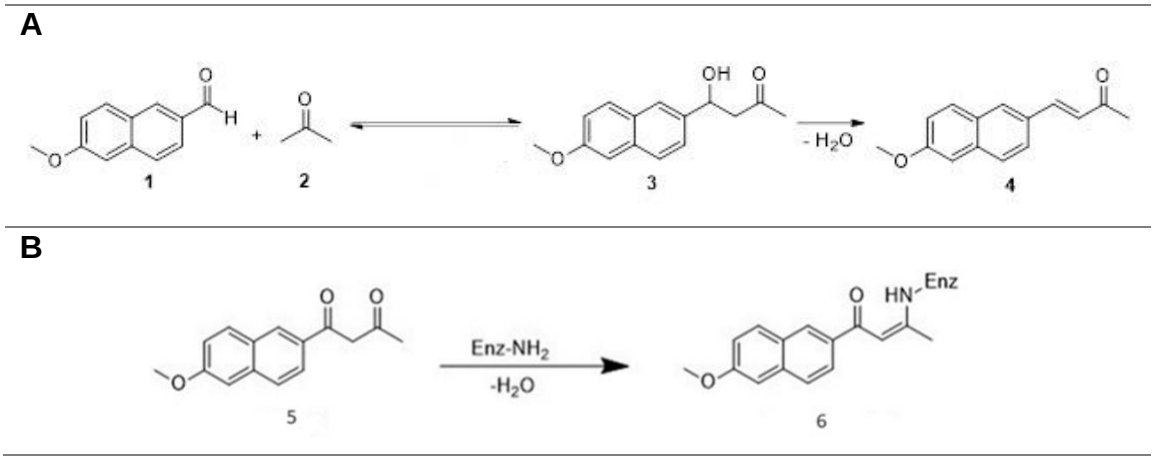

Figure 3. A) Reversable Aldol reaction of 6-methoxy-2-naphthaldehyde **1** and acetone **2** producing 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (Methodol, **3**) and the crossbedding a,b-unsaturated ketone **4**. **B)** Diketone inhibition reaction of 1,3 Diketone **5** forming Vinylogous amide **6** covalently with the catalytically active lysine.

As a proxy to assess the ability of the original 6H5L design and its variants to catalyze retro-aldol reaction through an active nucleophilic lysine, we conducted an inhibition reaction (Figure 3B) using the corresponding 1,3-diketone derivative **5** (1-(6-methoxy-2-naphthalenyl)-1,3-butanedione) of Methodol **2**. This derivative is expected to covalently react exclusively with the catalytically active lysine residue (H_2N-Enz) through a Schiff-base intermediate, forming a stable vinylogous amide **6**. The reaction was monitored using a UV-Vis spectrometer.

After the reaction of 1,3-diketone inhibitor **5** (Figure 3B) with 6H5L (and its six single lysine variants, the resulting UV-Vis absorbance spectra show a new peak at 340 nm compared to the

free protein (Figure 4A). This indicates the successful formation of vinylogous amide via Schiff-base, underlining the capability of these proteins to catalyze a retro-aldol reaction. The intensity of the vinylogous amide absorbance peak increases with an increasing protein-to-diketone ratio (1:1, 5, and 10), indicating more diketone binding with higher concentration (Figure 4A). The vinylogous amide peak of 6H5L has almost double the intensity compared to its variants (Figure 5A), suggesting activity in both binding pockets, while only one pocket is active in the single lysine variants.

To exclude spurious activity from surface residues and confirm specific activity through the lysines situated in the central channel, we performed competition assays between the diketone inhibitor and DPH. The first competition assay showed a decrease in fluorescence intensity for the diketone-labeled 6H5L in comparison to the free 6H5L protein (Suppl Figure S5A). This decrease persisted even at higher diketone inhibitor concentrations. The decrease in fluorescence intensity provides conclusive evidence of diketone binding within the hydrophobic channel of 6H5L, effectively blocking the central channel and preventing the binding of DPH. The reverse way of the competition assay, involving the titration of diketone inhibitor into 6H5L previously saturated with DPH dye, demonstrated a progressive decrease in fluorescence intensity with the addition of more diketone inhibitor with an inhibition constant (Ki) value of 0.27 mM (Suppl Figure S5B). These findings elucidate the heightened binding affinity of covalently bound diketone with 6H5L, displacing noncovalently bound DPH molecules from the hydrophobic channel. Consequently, this displacement led to the observed decrease in fluorescence, confirming the binding of the diketone inhibitor to the lysine residues within the central channel.

To test for retro-aldolase activity, a well-established fluorescence-based method to measure the formation of a naphthaldehyde derivative (6-methoxy-2-naphthaldehyde) product **1** by using a racemic Methodol **3** as a substrate was used (Figure 4A)[30]. The original 6H5L design demonstrated noticeable retro-aldolase activity with a second-order rate constant, k_{cat}/K_M, of 1.96 M^{-1} min^{-1} (Figure 5B. and Table 1). This rate is comparable to some of the first *de novo* retro aldolases[31]. This underscores its potential as a promising retro aldolase candidate by further improvement. An important validation experiment for this is checking the pH dependence of the monitored reaction. We found that the activity of 6H5L decreased at lower pH levels in the reaction media but remained significantly higher compared to the knockout version (6H5L_6R), which lacks any lysine residues inside the two binding pockets. The knockout version (6H5L_6R) exhibited almost no activity at pH 6, whereas 6H5L remained visibly active under the same conditions (Suppl Figure S6). This illustrates the active nucleophilicity of the lysines in the hydrophobic channel while maintaining inactive protonated lysines on the surface at lower pH.

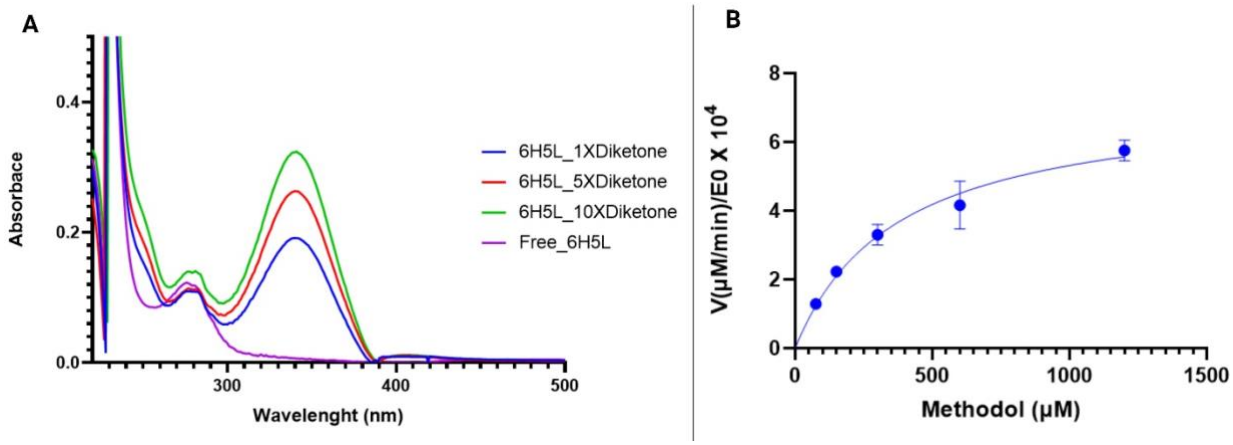

Figure 4. A) UV-VIS absorbance spectrum of a free 6H5L (27 µM) and labeled with varying diketone concentrations (1X, 5X, and 10X) resulting a vinylogous amide additional peak at 340 nm compared to the free protein. Higher diketone concentrations led to increased peak intensity. **B)** Retro-aldolase Michaelis–Menten plots with Methodol as a substrate of 6H5L design. E0, total enzyme concentration; V, initial reaction velocity.

Single lysine and knockout variants

To identify the most catalytically proficient lysine among the six within the hydrophobic channel, we designed six single-lysine variants (19K, 37K, 121K, 139K, 223K, 241K). Each variant features one lysine in the hydrophobic channel, with the other five lysine residues being replaced by arginines (Suppl Figure S3). Additionally, we used the knockout variant 6H5L_6R, which lacks any lysine residues within the two binding pockets as a negative control. All six single lysine variants exhibited measurable activity over time, albeit with lower activity than the original 6H5L (Figure. 5B and Table 1). This suggests the possibility of both cavities being active and highlights the influence proximity effects of nearby lysine residues in lowering the *pKa* of the nucleophilic lysine.

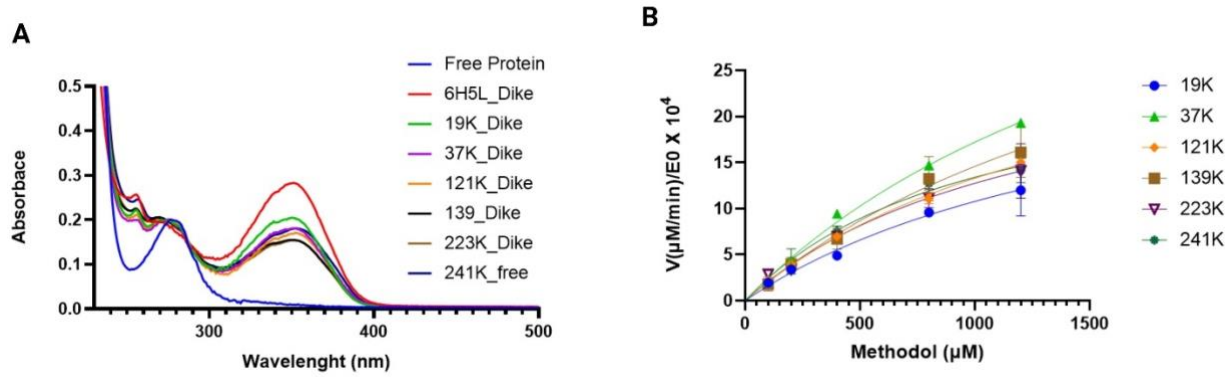

Figure 5. A) The UV-VIS absorbance spectrum of diketone-labeled 6H5L and all its single lysine variants (all at 27 µM) reacted with a concentration of 270 µM diketone revealed an additional vinylogous amide peak at 340 nm compared to the free protein. The vinylogous amide peak of 6H5L exhibited nearly double the intensity compared to its variants. **B)** Retro-aldolase Michaelis–Menten plot of all 6H5L single lysin variants with Methodol as a substrate. E0, total enzyme concentration; V, initial reaction velocity. Notably, all variants exhibited significant enzymatic activity.

Placing of a catalytic tetrad enhances the activity of the designed retro-aldolase

We aimed to enhance retro-aldolase's efficiency by designing an active site and altering the binding cavity for specificity towards the Methodol **3** substrate. This involves adding elements from the efficient RA95.5-8F [32] retro-aldolase into our helix bundle, specifically by incorporating two tyrosines, one glutamine, and a catalytic lysine (KYQY tetrad) to form a hydrogen bond network. Initially, we selected the single lysine variant 223K as our starting point and used Foldit as an interface to Rosetta to mutate three residues near this lysine to the corresponding tyrosines and glutamine. The resulting model was repacked and minimized. This process resulted in the design of the 223K_YQY (Figure 6A). In a second approach, we used RosettaMatch[33] for a comprehensive search for geometric matches of 1,3-diketone **5** (1-(6-methoxy-2-naphthalenyl)-1,3-butanedione) and the residues form the evolved active site of RA95.5-8F within our parent 6H5L design model. Initially, we confined this search to the central lysine resides and their immediate surroundings, which led to the design 37K_YQY. Furthermore, matching across the entire scaffold, led to the design variant 6H5L_199K (Figure 6A). Combinatorial sequence design in Rosetta[34] was used to optimize the amino acids surrounding the cavities hosting the ligand. The best scoring design was superimposed onto the best-ranked AF2 predicted structure of the same sequence.[2,3] Both 37K_YQY ad 6H5L_199K (Figure 6A) designs with the lowest rmsd were selected and ordered as synthetic genes for the purpose of evaluating the resulting protein folding behavior and catalytic efficiency. All 6H5L active site redesign versions were successfully produced and characterized using our standard methods (Supplementary Table S6). Both the rational (223K_YQY) and computational (37K_YQY) redesigns of the active site exhibited almost a 3-fold and 3.5-fold increase in catalytic efficiency compared to the single lysine variant 223K and the original 6H5L design, respectively (Suppl Figure. S7A&B and Table 1). To confirm that the increase in catalytic efficiency arises from the newly installed active site and redesigned cavity, knockout variants K223A_YQY and K37A_YQY were tested. The results showed a dramatic decrease in activity, reverting almost to the levels observed in the initial input designs (Suppl Figure S7C&D and Table S4). Notably, the 6H5L_199K version exhibited a remarkable 10-fold activity improvement achieved by repositioning the active site higher above the hydrophobic channel (Figure. 6A, 6B and Table 1). This strategic adjustment significantly enhanced the catalytic efficiency nearly twofold in the second-order rate constant (k_{cat}/K_M, of 20.95 M^{-1} min^{-1}) compared to the TIM-barrel computationally designed retro-aldolase RA95 (k_{cat}/K_M, of 11 M^{-1} min^{-1})[35] and slightly higher than RAβb-9 *de novo* β-barrel scaffold (k_{cat}/K_M, of 17 M^{-1} min^{-1})[36], addressing one of the most challenging aspects in enzyme evolution[37].

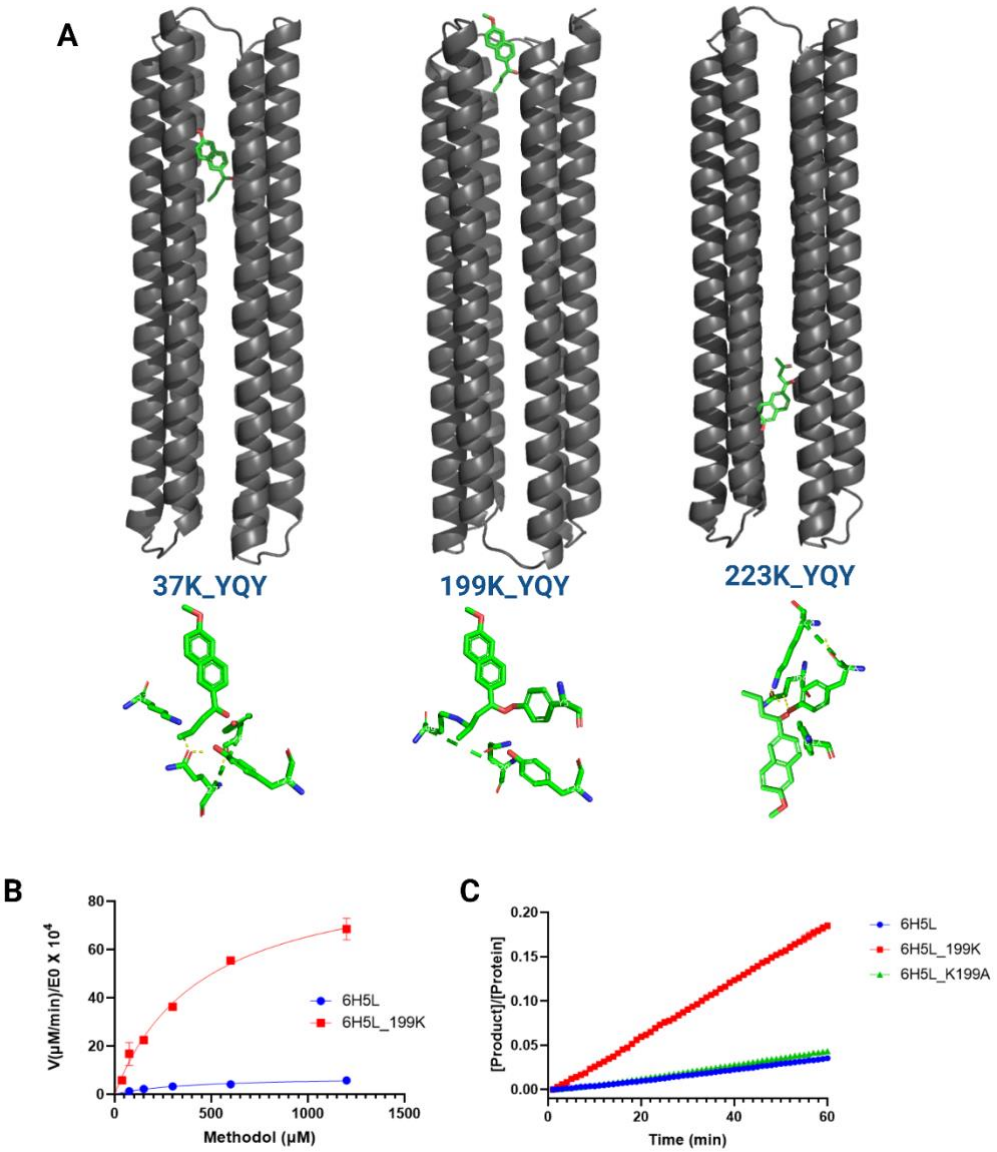

Figure 6. A) 6H5L active site redesign variants, including Methodol **3** substrate, are positioned in the redesigned pocket at different locations within the hydrophobic channel. **B)** Retro-aldolase Michaelis–Menten plots with Methodol as a substrate of 6H5L_199K (in red) active site redesigned variant compared to the initial 6H5L (in blue) activity of the input designs. E0, total enzyme concentration; V, initial reaction velocity. 6H5L_199K variant: remarkable 10-fold increase compared to 6H5L (Table 3). **C)** Retro-aldolase progress curves of 6H5L_199K (in red) active site redesigned variant of 6H5L compared to and its initial 6H5L (in blue) input design and 6H5L_K199A (in green) knockout version.

To confirm that the increase in catalytic efficiency originates from the expected lysine at position 199, the knockout version was designed and tested. By introducing an alanine at this position (6H5L_K199A), which displayed an activity level almost identical to the original 6H5L input design (Figure 6C and Suppl Table S4).

Based on the biophysical characterization and activity results, it's evident that the targeted changes to the initial 6H5L backbone/sequence, especially within the hydrophobic channel, have

313 led to improved activity without significantly influencing the overall structure (Supplementary Table
314 S6).

315 **Table 1.** Kinetic parameters of retro-aldolase for active site redesigned variants of 6H5L and its initial input designs, in
316 comparison to a variety of initially computationally designed enzymes (assessed using the racemic Methodol).

Protein	K_m (μM)	k_{cat} (x 10^{-4} min^{-1})	k_{cat}/K_M (M^{-1} min^{-1})
6H5L	371.4±61.84	7.29±0.07	1.96
223K	967 ±211	9.5 ±1.1	0.98
223K_YQY	1550 ±225	45.0 ±4.2	2.90
37K_YQY	1037 ±128	74.00 ±5.27	7.14
6H5L_199K	453.2 ±49.31	94.95±4.54	20.95
RA95 [35]	530	60	11
RAβb-9 [36]	--	--	17

317

318 **A practically useful de novo enzyme - whole cell biotransformations**

319 To be applicable on industrial scales, most biocatalysts are used in a whole cell fashion, meaning
320 that they could be produced in large fermenters and used directly in reaction mixtures. To
321 showcase the high versatility of our designs, we tested 6H5L and its highest active variant
322 6H5L_199K as whole cell biocatalysts for the aldol reaction (Figure 3A) of 6-methoxy-2-
323 naphthaldehyde **1** (5 mM) and acetone **2** (5% v/v). Both designs exhibited significant activity, with
324 substrate consumption around 36% and 46%, respectively, while enzyme-free cells showed only
325 16% conversion (Figure 7). The aldol product was found to undergo dehydration to the
326 corresponding α,β-unsaturated ketone **4**, and control reactions indicated that this condensation
327 was in the most part catalyzed by the protein (*vide infra*).

328 Additionally, 6H5L overexpressing cells catalyzing the retro-aldol reaction (Figure 3A) on 5 mM
329 racemic Methodol as substrate (starting compound **3** used for the retro-aldol reaction showed a
330 slight enantiomeric excess of 30%), showed comparable consumption of both enantiomers (18%
331 and 17%, respectively), indicating a non-enantioselective behavior. A similar trend was observed
332 with 6H5L_199K cells (19% and 17% consumption for each enantiomer, respectively). The
333 enzyme-free cells exhibited lower substrate consumption, with 9% for each enantiomer,
334 respectively (Suppl Figure S8). Notably, the catalyst preparation was compatible with the organic
335 solvent acetonitrile, up to 20 vol%.

In line with the observed formation of the aldol condensation product through dehydration as major product during the reaction between **1** and **2**, we could detect the formation of **4** during the retro-aldol reaction with **3** and confirmed that this reaction was catalyzed by the cells. Demonstrating significant aldolase and retro-aldolase activity in whole cells, both from the original 6H5L design and the 6H5L_199K active site redesign variant, highlights the promising potential of the parent non-natural fold to be used as a highly engineerable reaction chamber. Combined with its high thermodynamic stability and solvent tolerance to 20 vol% this design established itself as a promising candidate for optimization and utilization in green biotechnology applications.

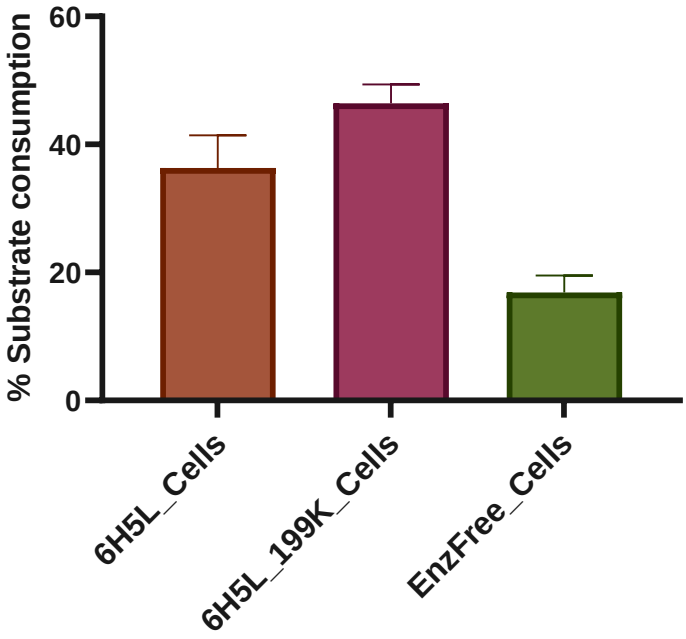

Figure 7. Substrate consumption in aldol reaction catalyzed by 6H5L and 6H5L_199K variants (whole cells) compared to enzyme free cells (EnzFree_Cells) using 5 mM 6-methoxy-2-naphthaldehyde **1** and 5 vol% acetone (**2**) in 50 mM sodium phosphate buffer (pH 7.5) containing 150 mM NaCl and 10 vol% acetonitrile.

Conclusions and Outlook

Enzyme design efforts currently rely on designing and screening dozens to hundreds of hypothetical enzymes to produce a viable catalyst. While this is a highly efficient strategy, it lacks the ease that many of the orthogonal approaches, like organocatalysis, exhibit - to quickly and rationally engineer and repurpose catalysts for envisioned reactions. One way of gaining quicker access to products of interest and using the exquisite asymmetric environment that a protein provides, is via a highly versatile and stable protein chassis with a high mutational tolerance. Our study constitutes a significant step forward in this direction. Using the highly accurate backbone

generation via parametric design, we demonstrated that such a scaffold for holding catalytic centers in an asymmetric environment can be made and yields moderately efficient enzymes while screening as little as nine variants. While the most active design does not show an improvement over previous in-silico designs, its high thermostability and solvent tolerance allows it to be used as an efficient whole cell biocatalyst. This highlights the compromise between ease of engineering and high-rate accelerations, usually only achieved through lengthy laboratory evolution campaigns. However, the ability to scaffold catalytic centers in a robust protein chassis with high accuracy now allows the type of experiments that can elucidate such structure- and function relationships quickly, helping to build models for synthetic strategies in biomolecules.

Thorough biochemical characterization allowed us to identify practically useful enzymes that exhibited high selectivities and moderate turnover in whole cell biotransformations. For generalized models of proteins being used as asymmetric reaction chambers, datasets like this will be needed for a diverse set of chemical transformations. In addition, the ability to design protein scaffolds like ours could foster the creation of a set of asymmetric environments that can be highly engineered to be used as versatile biocatalysts. Much like the streptavidin system has been used to anchor artificial cofactors into an asymmetric environment and thereby fostered the development of new synthesis routes, our approach could drive the development of methods for reusability of de novo designed functional proteins.

Finally, whole cell biotransformations have played a pivotal role in biocatalysis campaigns to create synthetic routes to new pharmaceuticals and complex fine chemicals. Here, we established a protein scaffold which shows great promise in being generally applicable for biotransformations.

Methods

Protein production and purification:

All individual protein variants were ordered as gene fragments and cloned using Gibson assembly[38] into the pEThTEV plasmid (Suppl Figure. S1), which includes an N-terminal His-Tag with a TEV cleavage site. The successfully confirmed cloned genes were transformed into *E. coli* BL21 Star (DE3) competent cells for protein production. Single colonies were used to inoculate 20 mL LB pre-culture medium containing 50 mg/L Kanamycin and incubated overnight at 37 °C and 180 rpm. One liter of TB main culture containing 100 mg/L Kanamycin was inoculated by adding 10 mL of pre-culture and incubated at 37 °C,180 rpm until an OD600 around 0.6 was reached. For protein induction, 1 mM IPTG was added to the main culture and incubated overnight at 18 °C and 180 rpm. The cells were harvested by centrifugation at 4 °C / 4500 rpm for 30 min and stored at -20 °C before lysis. The cells were lysed by sonication for 10 min under cooling in

lysis buffer (20 mM Na_3PO_4 pH 7.4, 500 mM NaCl, 10 mM imidazole, and 10 mM PMSF). The lysate product was centrifuged at 18,000 rpm for 45 min to isolate the cell debris and the insoluble protein.

The protein variants were purified by Ni-NTA affinity chromatography and size-exclusion chromatography (SEC), including His-Tag cleavage using TEV protease for further biochemical characterization as follows. The cleared lysate was loaded onto the lysis buffer-equilibrated Ni-NTA column and washed with a wash buffer containing the same buffer but with 30 mM imidazole. The target protein was eluted from the column using elution buffer (20 mM Na_3PO_4 pH 7.4, 500 mM NaCl, 250 mM imidazole), concentrated and buffer exchanged to 20 mM Tris-HCl, 150 mM NaCl, pH 8. Tag cleavage was performed by adding 15 mM TEV protease to the solution at a protein concentration of 1 mg/mL and incubating overnight at room temperature. The cleaved protein was separated from the mixture by loading it again onto a Ni-NTA column and collecting the flowthrough. Size-exclusion chromatography was performed as the second purification step, using Xtal buffer as the mobile phase over Superdex 200 column. SDS-PAGE gels were run to monitor the protein throughout all purification steps. The concentrations of the pure proteins were calculated by measuring the absorbance of the samples at a wavelength of 280 nm using the A280 measurement option of the NanoDrop™ 2000 for all further biochemical characterizations.

Circular dichroism spectroscopy (CD):

For CD measurements, the pure proteins were run over Size-Exclusion Chromatography (SEC) once again using CD buffer (20 mM sodium phosphate and 150 mM sodium fluoride at pH 8.0) as the mobile phase. The eluted fractions were collected and concentrated to 1 mg/ml. Measurements were carried out using quartz capillaries (Helix Biomedical Accessories, Inc.) with a path length of 0.5 mm on a CD spectrometer (J-1500, JASCO, USA). Wavelength scan spectra were measured from 260 to 190 nm, and temperature scans were performed with incremental steps of 2°C from 20°C to 95°C and back to 20°C. The CD signals were corrected using a blank sample containing the same CD buffer composition.

Protein solvent stability:

The purified 6H5L protein was dissolved in a buffer of 20 mM Tris-HCl and 150 mM NaCl, pH 8, to a final concentration of 1 mg/ml, with DMSO concentrations varying from 0% to 50%. Before each measurement, the mixture was centrifuged for 5 minutes at 14,500 rpm and 4°C for clarification. UV-Vis absorbance spectra and protein concentrations were measured at different time intervals (0, 3, and 24 hours) to monitor the protein's solubility under each condition.

Ligand-binding

Ligand-binding assays were performed using 1,6-diphenylhexatriene (DPH) or 1-(4-Trimethylammoniumphenyl)-6-Phenyl-1,3,5-Hexatriene p-Toluenesulfonate (TMA-DPH) fluorescence days and all our designed helix bundle variance, The experiments maintained a constant ligand concentration of 1 µM, while the designed proteins concentrations varied from 0.78 or 1.6 up to 50 or 100 µM in 20 mM Tris, 150 mM NaCl, 1% DMSO, pH 8 buffer at 25 °C. After a 1-hour incubation with shaking at 300 rpm, fluorescence spectra were obtained from the top of the samples in a 96-well plate, using specific excitation and emission wavelengths (Table 5). The dissociation constant (K_D) values were determined by fitting the data to a one-site_total binding model using GraphPad Prism 8. All the measurements were carried out in 96-Well Black Flat Bottom Polystyrene Non-Binding Surface Microplates (Corning, Lowell, MA) using a BMG Labtech CLARIOstar PLUS Microplate Reader (Aylesbury, UK).

Small-angle X-ray scattering (SAXS):

For SAXS measurements, each protein variant reloaded over Size-Exclusion Chromatography (SEC) Superdex 200 Increase 10/300 GL column in a SAXS buffer (20 mM Tris-HCl, 150 mM NaCl, 2 mM TCEP, 3% glycerol, pH 8). The resulting pure protein fractions were collected, concentrated, and subsequently adjusted to a concentration of 3 mg/ml, calculated using the formula *C (mg/ml) = 100 / MW (kDa)*. The samples were measured at ESRF, BM29 BioSAXS of Grenoble (France). All experimental scattering data were corrected by subtracting the buffer scattering data and then fitted to the theoretically calculated curve of AlphaFold2 predicted models[25,28] using the *FoXS* online server to calculate the quality of fit, chi² values[29].

X-Ray crystal structure

To improve the quality of the protein crystals by slightly decreasing the protein solubility, we redesigned the 6H5L surface by replacing 15 lysine residues on the equivalent helical wheel *f* position with glutamine (6H5L_FKQ variant), resulting in a decreased pI from 9 to 5.7[39,40]. The prediction of the 6H5L_FKQ variant was carried out using Alphafold2 and compared to the original 6H5L. 6H5L_FKQ variant was successfully produced in *E. coli* BL21 Star (DE3) competent cells and thoroughly characterized using our standard methods (Suppl Table S7).

Crystallization drops were set up with commercial crystallization screens using the vapor diffusion method employing a mosquito® Xtal3 crystallization robot (SPT Labtech) and incubated at 293 K. The protein concentration was 20 mg/ml in 20 mM Tris-HCl, 150 mM NaCl, 2 mM TCEP pH 8.0 and the drop volume was 400nL, with a 1:1.5 ratio of protein and precipitant solution. Crystallization drops were equilibrated against a reservoir containing 40 µL of precipitant solution. The crystals were obtained from manually set up drops of 1 µL protein mixed with 1.5 µL precipitant solution (17.5% PEG 3350, pH = 7.8, 150 mM L-Malic Acid. 80 µL of precipitant solution

in the reservoir). Crystals appeared after one day to two weeks. The obtained crystals were harvested from mother liquor with CryoLoops (Hampton Research) and briefly incubated with mother liquor containing 25% glycerol or 25% PEG400 followed by flash freezing in liquid nitrogen. Diffraction data was collected at 100 K on beamlines MASSIF-3 and ID30B at ESRF, Grenoble (France). Full datasets (360°) were collected to 2.8 Å resolution.

The collected data were processed using XDS[41]. with the provided input file from the beamline. Structure determination was performed by molecular replacement using PHASER[42] with the design models as search templates. The best solution was refined in reciprocal space with PHENIX[43] with 5% of the data used for R_{free} and by real-space fitting steps against σA-weighted $2F_o–F_c$ and $F_o–F_c$ electron density maps using COOT[44] . Water molecules were placed automatically into difference electron density maps and accepted or rejected according to geometry criteria and their B-factors, defined in the refinement protocol. Details of data collection, processing, and refinement are summarized in Supplementary Table S7.

Inhibition reaction:

The diketone inhibition reactions (Figure 3B) for all protein variants were carried out at 29 °C in a buffer solution (50 mM sodium phosphate, 150 mM NaCl, and 5% DMSO) at pH 8. The reaction mixture consisted of a protein concentration of 28 μM and 28-280 μM 1,3-diketone **5** (1-(6-methoxy-2-naphthalenyl)-1,3-butanedione) (1:1-10 ratio) in a total reaction volume of 1 mL. After an overnight incubation period, the reaction mixture was loaded over SEC to remove the excess diketone, collecting only the protein fraction to measure. For a clear comparison, UV-Vis spectra were obtained for the free protein as well as the labeled protein reaction product.

Competition assay:

We have implemented two distinct competition assays. The first assay (DPH_diketone competition assay) involved carrying out a DPH fluorescent binding experiment, where we compared the fluorescence intensities between the free protein and the proteins labeled with diketone inhibitor. To ensure consistent protein concentrations, we standardized all the SEC fractions of the treated proteins to possess identical absorbance values at 280nm before initiating the fluorescence measurements (Figure. 4A & 5A).

The second experiment included a titration assay, where diketone was incrementally added (up to 1.72 mM) to the saturated protein-DPH complex (1 μM of each after 1 hour of incubation with shaking at 300 rpm), and the corresponding fluorescent signals for DPH dye (Supplementary Table S2) were observed at various titration points.

Both measurements were carried out in 96-Well Black Flat Bottom Polystyrene Non-Binding Surface Microplates (Corning, Lowell, MA) using a BMG Labtech instrument (Aylesbury, UK) Clariostar plate reader. To ensure accuracy and account for background effects, fluorescence signals were measured for both the buffer-DPH and buffer-DPH-diketone mixtures under identical conditions.

Retro-aldolase activity:

Kinetic assays of all protein variants were conducted in 96-Well Black Flat Bottom Polystyrene Non-Binding Surface Microplates (Corning, Lowell, MA) using a plate reader (CLARIOstar, BMG Labtech). Initial reaction rates for the conversion of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone **3** (Methodol) substrate to 6-methoxy-2-naphthaldehyde **1** (Figure 3A) were determined by monitoring the formation of the fluorescent product with an excitation wavelength *λex* 330 nm and emission wavelength *λem* 452 nm measured from the top (with a bandwidth of 2 nm)[30]. All measurements were conducted at 29 °C in assay buffer (50 mM sodium phosphate, 150 mM NaCl, and 7.5% DMSO) at pH 8. The stock of Methodol substrate was dissolved in DMSO and stored at -20 °C. The fluorescence signal was corrected for the buffer-catalyzed background reaction under the same conditions and converted to product concentration using a calibration curve within the product formation range. The data were subjected to Michaelis–Menten fitting using GraphPad Prism 8 to determine catalytic rate constant (K_{cat}) and Michaelis constant (K_M) values.

Whole cells biotransformation:

E. coli BL21 Star (DE3) cells were lyophilized after overexpressing either 6H5L or 6H5L_199K and their catalytic activity was tested in aldol and retro-aldol reactions. The same kind of cells with enzyme empty plasmid were used as control. The reaction setup was as follows: 15 mg of lyophilized cells were resuspended in the reaction buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.5) for 20 min at 30 °C under shaking at 120 rpm. Subsequently, 5 mM of substrate was added to each reaction. In the aldol reaction, 6-methoxy-2-naphthaldehyde **1** was added from a stock solution in acetonitrile (10 vol% final concentration co-solvent) followed by 5 vol% acetone. In the retro-aldol reaction, 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone **3** (Methodol) was added from a stock solution in acetonitrile (20 vol% final concentration co-solvent). The reactions (final volume 500 µL) were incubated overnight at 30 °C and 120 rpm. Control reactions were also run in the absence of cells.

The products were extracted using ethyl acetate (2 x 250 µL) spiked with 0.5 vol% acetophenone as internal standard (IS). The combined extracts were dried over anhydrous sodium sulfate and

analyzed by HPLC using an injection volume of 5 µL and a Chiralpak® AS-H column (250 x 4.6 mm, 5 µM, Daicel chiral technologies) for the aldol reaction and a Chiralcel® OD-H column (250 x 4.6 mm, 5 µM, Daicel chiral technologies) for the retro-aldol reaction. The identity of the compounds was further confirmed by GC/MS. The results were corrected for the buffer-catalyzed background reaction under the same conditions.

ACKNOWLEDGEMENTS

We thank David Baker for detailed discussion in the beginning of this project and are grateful for his sustained advice throughout. We also want to thank Dek Woolfson for inspiring conversations about surface residue compositions in coiled coils and their crystallizability. Additionally, we thank Peter Macheroux for important remarks about the kinetics of the designed system and Lukas Schober for reproducing the synthesis method for Methodol. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities, and we would like to thank the staff of the ESRF and EMBL Grenoble for assistance and support in using beamlines MASSIF-3, ID30B, and BM29. W.E. and G.O. were supported by the Austrian Science Fund (FWF) grant 10.55776/P30826 and by funding from the European Research Council through an ERC Starting Grant (HelixMold 802217). H.L. and G.O. were supported by a FETOPEN project (ARTIBLED, 863170).

Author contributions: Conceptualization: W.E., G.O.; Methodology: W.E, M.H, G.O.; Validation: W.E., B.G., D.S., H.L., M.H., G.O.; Formal analysis: W.E., M.H., G.O.; Investigation: W.E, B.G., D.S., M.C., H.L., M.A., M.H., G.O.; Resources: M.H., G.O.; Data Curation: W.E., G.O.; Writing: W.E., G.O.; Visualization: W.E., G.O; Supervision: M.H., G.O.; Project administration: G.O.; Funding acquisition: M.H., G.O.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: Custom Python scripts, RosettaScripts and command lines, will be made available upon peer-reviewed publication.

References

1. Hanefeld, U., Hollmann, F. & Paul, C. E. Biocatalysis making waves in organic chemistry. *Chem Soc Rev* **51**, 594–627 (2022).

2. Bell, E. L. *et al.* Biocatalysis. *Nature Reviews Methods Primers* **1**, 46 (2021).

3. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie International Edition* **57**, 4143–4148 (2018).

4. Cui, H. *et al.* Less Unfavorable Salt Bridges on the Enzyme Surface Result in More Organic Cosolvent Resistance. *Angewandte Chemie International Edition* **60**, 11448–11456 (2021).

5. Buller, R. *et al.* From nature to industry: Harnessing enzymes for biocatalysis. *Science (1979)* **382**, (2023).

6. France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* **3**, 715–735 (2023).

7. Xie, W. J. & Warshel, A. Natural Evolution Provides Strong Hints about Laboratory Evolution of Designer Enzymes. *Proceedings of the National Academy of Sciences* **119**, (2022).

8. Kalvet, I. *et al.* Design of Heme Enzymes with a Tunable Substrate Binding Pocket Adjacent to an Open Metal Coordination Site. *J Am Chem Soc* **145**, 14307–14315 (2023).

9. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).

10. Mansoor, S., Baek, M., Juergens, D., Watson, J. L. & Baker, D. Zero-shot mutation effect prediction on protein stability and function using <scp>RoseTTAFold</scp>. *Protein Science* **32**, (2023).

11. Albanese, K. I. *et al.* Rationally seeded computational protein design of α-helical barrels. *Nat Chem Biol* **20**, 991–999 (2024).

12. Yeh, A. H.-W. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).

13. Kalvet, I. *et al.* Design of Heme Enzymes with a Tunable Substrate Binding Pocket Adjacent to an Open Metal Coordination Site. *J Am Chem Soc* **145**, 14307–14315 (2023).

14. Lipsh-Sokolik, R. *et al.* Combinatorial assembly and design of enzymes. *Science (1979)* **379**, 195–201 (2023).

15. Braun, M. *et al.* Computational design of highly active de novo enzymes. *bioRxiv* 2024.08.02.606416 (2024) doi:10.1101/2024.08.02.606416.

16. Lauko, A. *et al.* Computational design of serine hydrolases. *bioRxiv* 2024.08.29.610411 (2024) doi:10.1101/2024.08.29.610411.

17. Tsuboyama, K. *et al.* Mega-scale experimental analysis of protein folding stability in biology and design. *Nature* **620**, 434–444 (2023).

18. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science (1979)* **357**, 168–175 (2017).

593 19. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles.
594 *Science (1979)* **346**, 481–485 (2014).

595 20. Benke, S. *et al.* The assembly dynamics of the cytolytic pore toxin ClyA. *Nat Commun* **6**, 6198
596 (2015).

597 21. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles.
598 *Science (1979)* **346**, 481–485 (2014).

599 22. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and
600 frameworks. *Nat Methods* **17**, 665–680 (2020).

601 23. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design.
602 *J Chem Theory Comput* **13**, 3031–3048 (2017).

603 24. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
604 (2021).

605 25. Guo, H.-B. *et al.* AlphaFold2 models indicate that protein sequence determines both structure and
606 dynamics. *Sci Rep* **12**, 10696 (2022).

607 26. Albanese, K. I. *et al.* Rationally seeded computational protein design of α-helical barrels. *Nat Chem*
608 *Biol* **20**, 991–999 (2024).

609 27. Thomas, F. *et al. De Novo* -Designed α-Helical Barrels as Receptors for Small Molecules. *ACS Synth*
610 *Biol* **7**, 1808–1816 (2018).

611 28. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682
612 (2022).

613 29. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS Profile Computation
614 and its Assessment by Contrast Variation Experiments. *Biophys J* **105**, 962–974 (2013).

615 30. Tanaka, F., Fuller, R., Shim, H., Lerner, R. A. & Barbas, C. F. Evolution of Aldolase Antibodies in
616 Vitro : Correlation of Catalytic Activity and Reaction-based Selection. *J Mol Biol* **335**, 1007–1018
617 (2004).

618 31. Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science (1979)* **319**, 1387–
619 1391 (2008).

620 32. Obexer, R. *et al.* Emergence of a catalytic tetrad during evolution of a highly active artificial
621 aldolase. *Nat Chem* **9**, 50–56 (2017).

622 33. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*
623 *(1979)* **302**, 1364–1368 (2003).

624 34. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design.
625 *Protein Science* **15**, 2785–2794 (2006).

626 35. Althoff, E. A. *et al.* Robust design and optimization of retroaldol enzymes. *Protein Science* **21**, 717–
627 726 (2012).

628 36. Kipnis, Y. *et al.* Design and optimization of enzymatic activity in a de novo β-barrel scaffold. *Protein*
629 *Science* **31**, (2022).

630 37. Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic
631 efficiency. *Biochemistry* **15**, 5631–5640 (1976).

632 38. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat*
633 *Methods* **6**, 343–345 (2009).

634 39. Anstrom, D. M., Colip, L., Moshofsky, B., Hatcher, E. & Remington, S. J. Systematic replacement of
635 lysine with glutamine and alanine in *Escherichia coli* malate synthase G: effect on crystallization.
636 *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 1069–1074 (2005).

637 40. Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. A Set of *de Novo* Designed Parallel
638 Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-
639 nanomolar Regime. *J Am Chem Soc* **135**, 5161–5166 (2013).

640 41. Kabsch, W. *XDS*. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).

641 42. McCoy, A. J. *et al. Phaser* crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).

642 43. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
643 electrons: recent developments in *Phenix*. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).

644 44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta*
645 *Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

646