Computational design of highly active *de novo* enzymes

Markus Braun*[1], Adrian Tripp*[1], Morakot Chakatok[1], Sigrid Kaltenbrunner[1], Massimo Totaro[1], David Stoll[1], Aleksandar Bijelic[1], Wael Elaily[1], Shlomo Yakir Hoch[1,2], Matteo Aleotti[3], Mélanie Hall[3,4] and Gustav Oberdorfer[1,5,#]

[1]Institute of Biochemistry, Graz University of Technology, Petersgasse 12/2, 8010-Graz, Austria
[2]Weizmann Institute of Science, Rhovolt, Israel
[3]Institute of Chemistry, University of Graz, Heinrichstrasse 28, 8010 Graz
[4]BioHealth, Field of Excellence, University of Graz, Austria
[5]BioTechMed Graz, Mozartgasse 12/2, 8010-Graz, Austria
* These authors contributed equally #correspondence to: gustav.oberdorfer@tugraz.at

Abstract

Custom designed enzymes can further enhance the use of biocatalysts in industrial biotransformations, thereby helping to tackle biotechnological challenges of the 21st century. We present rotamer inverted fragment finder – diffusion (Riff-Diff), a hybrid machine learning and atomistic modeling strategy for scaffolding catalytic arrays in *de novo* protein backbones with custom substrate pockets. We used Riff-Diff to scaffold a catalytic tetrad capable of efficiently catalyzing the retro-aldol reaction. Functional designs exhibit a high fold diversity, with substrate pockets similar to natural enzymes. Some of the thus generated *de novo* enzymes show activities rivaling those optimized by in-vitro evolution. The design strategy can, in principle, be applied to any catalytically competent amino acid constellation. These findings are paving the way to address factors for the practical applicability of *de novo* protein catalysts in industrial processes and shed light on fundamental principles of enzyme catalysis.

Introduction

Natural enzymes have emerged as an indispensable tool for chemists, offering unparalleled precision and selectivity in chemical transformations (*1*). However, identification of natural enzymes with the desired activity can require extensive amounts of resources and screening capabilities. Very recently, protein design paved the way for the creation of biocatalysts designed for arbitrary functions (*2–5*). Yet, an often underemphasized detail of these studies were the low initial catalytic rates of the designed enzymes. As a result, the current paradigm to design efficient biocatalysts is to compensate for low initial efficiencies with high-throughput screening and directed evolution (*6–8*). While this approach is valuable and robust, it is not suited to create novel, proficient catalysts for chemical reactions inaccessible through high-throughput screening. For this reason, it is necessary to advance our ability to design efficient enzymes zero-shot. This, however, requires addressing current methodological shortcomings in functional design and prediction of enzymatic activity which will ultimately deepen our understanding of the fundamental principles of enzyme catalysis.

De novo enzyme design strategies are commonly based on the transition-state model of enzyme catalysis (*9*). The model proposes that functional groups in the active site accelerate chemical reactions by stabilizing the reaction's transition state over the ground state. In accordance with this model, minimal active sites, called theozymes, can be constructed by placing amino acid functional groups around a computational model of a transition state in a stabilizing geometry (*10*). In the first computationally designed enzymes, these theozymes were introduced into cavities of natural proteins, which could support the desired geometry (*11–13*). The thus designed enzymes successfully catalyzed chemical transformations, some of which considered new to nature at the time, but with low catalytic rate accelerations and often low success-rates (*14–22*). Unsurprisingly, among the main limiting factors of the approach was the accuracy with which the catalytic side chains could be placed in the active site. Successful designs exhibited less than 1 Å side chain RMSD between the geometry in the theozyme and a high resolution experimental structure. Inactive designs on the other hand, showed unproductive conformations (*23–25*). Additionally, scaffolds of natural proteins often did not tolerate the mutations introduced and thus could not be solubly expressed. The recent transformative successes in accurate structure prediction methods and protein design tools warrants another attempt to solve this problem (*26–28*).

Since the first successful attempts, critical analysis of designed enzymes and their evolved variants revealed additional aspects to the enzyme design problem. One of them was the construction of catalytically potent theozymes (*24*, *29*, *30*). This was demonstrated by early

mutational studies of designed retro-aldolases, for which multiple catalytic interactions were programmed to stabilize the reaction's transition states (*13*). Of the programmed interactions, only the catalytic lysine in a hydrophobic binding pocket significantly contributed to catalysis. Directed evolution of one of these designs even replaced the initial catalytic machinery with a completely new catalytic tetrad in the final variant. The geometry of this experimentally evolved tetrad was unlike any of the calculated theozymes, which further hinted at the low quality of initial theozymes (*31*). Furthermore, molecular dynamics simulations of retro-aldolases at intermediate steps during directed evolution demonstrated the importance of preorganization in catalytic arrays (*32*). The conformational populations of catalytic amino acids were increasingly rigidified towards catalytically productive conformations. Moreover, conformational distributions of active site residues present in kemp-eliminases, were shown to discriminate between active and inactive designs (*33*–*35*). In addition, backbone to side chain compatibility, as well as atomic density inside an active site pocket were found to correlate strongly with high activity for a class of recombined de novo designed xylanases (*36*). Despite all of these findings, recent *de novo* enzyme design strategies rarely optimized preorganization explicitly.

In this study, we investigated whether efficient biocatalysts can be created *zero-shot* by incorporating an optimized catalytic center in *de novo* protein backbones. We developed a method called Riff-Diff (**R**otamer **i**nverted **f**ragment **f**inder – **Diff**usion) which focuses on precise placement of idealized backbone-side chain fragments, preorganization of given catalytic arrays and the design of backbones with custom binding pockets. To test the effectiveness of our approach, we designed retro-aldol enzymes based on a previously reported, highly optimized catalytic tetrad (*31*). Detailed biochemical, biophysical and structural characterization, as well as molecular dynamics simulations allowed us to test current hypotheses about precision of placement of active site atoms and their preorganization, backbone - side chain compatibility and conformational dynamics in enzymes. The extraordinary wealth of this data comes from the exact same catalytic environment placed into different backbones, allowing us to draw conclusions for *de novo* enzyme design that were not obvious before.

Results

Artificial motif libraries scaffold catalytic arrays

Previous work in *de novo* enzyme design has established the importance of potent catalytic arrays for high catalytic rate accelerations (*2*, *4*, *23*, *37*). The conformational preorganization of such catalytic arrays was found to further promote activity (*32*, *34*, *38*). To promote preorganization during design, we embed the individual amino acids of a catalytic array into idealized alpha-helical fragments. The rotamers of each amino acid are selected to be compatible with the phi-psi angle combination of the fragment backbone, thus ensuring the best possible backbone-side chain combination (*39*). In addition, this ensures minimal energetic penalty is associated with the preferential catalytic rotamer. We then combine the individual fragments into *artificial* motifs which can be scaffolded by motif-scaffolding tools like RFdiffusion, Chroma, or others (*26*, *40–43*).

Success rates for motif-scaffolding with RFdiffusion have been shown to correlate with increasing size of the selected motifs. Our artificial motifs are relatively large but they are combined in arbitrary geometries based on the positions of the backbone atoms of the catalytic array. The arbitrary geometry can result in physically implausible motifs, leading to low scaffolding success rates. To increase scaffolding success, we developed a method that generates and assembles libraries of artificial motifs from catalytic arrays, from which high quality motifs can be sampled. The method starts by *inverting* rotamers of the catalytic array, which fixes the functional groups in space and varies the position of their backbone atoms. Rotamers are selected for backbone compatibility with the idealized helical fragments which are subsequently placed onto the rotamer's backbone atoms. Next, a search for all possible combinations of fragments that are not sterically clashing with each other is performed. The identified assemblies are aggregated and stored in an artificial motif library. Non-clashing assemblies are ranked according to the rotamer probabilities of the individual active site residues (Figure 1A).

Ensuring accessible and highly optimized active site pockets

An essential feature of enzymes as biocatalysts is their high substrate-specificity which is facilitated by well defined interactions of the substrate with the enzyme's asymmetric binding pocket. To analyze the extent to which substrates are typically buried within binding pockets, we extracted an enzyme dataset from the PDBbind database (*44*). In this dataset, an average of

7.1 α-carbons were within 8 Å of the substrate, normalized by the number of ligand heavy atoms (Figure 1C). For enzyme design, vanilla RFdiffusion implements an auxiliary potential to mimic the physical presence of a substrate. We found that this potential decreased the number of clashes between ligand and backbone, but did not promote the formation of well defined substrate binding pockets, leading to a trade-off between clashes and substrate interactions (Figure 1C). For this reason, we added another alpha-helix as an entry-channel placeholder to each artificial motif at the position of the binding pocket prior to motif scaffolding. We then implemented a custom auxiliary potential, which places the center of the denoising trajectory on the helix and enforces a distance constraint of all backbone atoms to this center. The 'placeholder' helix is removed after diffusion, leaving a vacant binding pocket. This procedure allowed us to drop the substrate auxiliary potential and enabled diffusion of protein backbones with custom pockets.

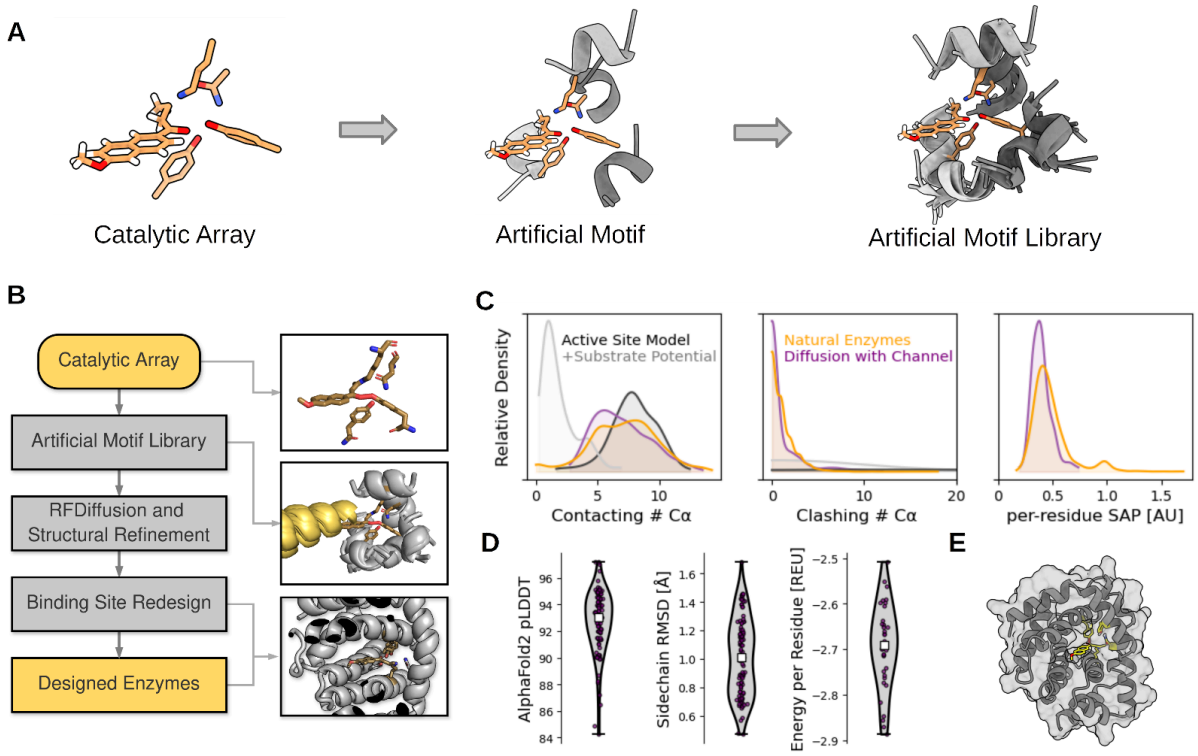

Figure 1: The Riff-Diff pipeline designs enzymes reminiscent of natural enzymes. A. Artificial motif libraries are collections of artificial motifs which are constructed from theozymes. **B.** Schematic overview of the semi-automated Riff-Diff pipeline. **C.** Binding pockets of natural enzymes (yellow) typically bury their substrates, measured by the number of alpha-carbons within 8 Å of the substrate. RFdiffusion's substrate potential offers only a trade-off between substrate burial and clashes (overlaps) with the pocket (light- and dark gray). Riff-Diff (purple) scaffolds enzyme backbones that bury the substrate in binding pockets reminiscent of natural enzymes. Surface aggregation propensity of the designed enzymes is similar to natural enzymes. **D.** Riff-Diff produces high-quality enzyme backbones - and sequences. **E.** An exemplary design shows the substrate (methodol) buried deep within the enzyme's active site (yellow).

Riff–Diff robustly scaffolds catalytic arrays

An integral part of Riff-Diff is its backbone refinement protocol, in which the diffused enzyme backbones are refined in an iterative fashion. Recent work showed that the quality of *de novo* backbone-side chain pairs improves when predicted structures are used for subsequent sequence design (*45*, *46*). Therefore, we implemented iterative refinement cycles into Riff-Diff during which the coordinates of idealized helical fragments of the artificial motifs were used as constraints to minimize the motif backbone RMSD. Additionally, Rosetta's FastDesign protocol (*47*) explicitly optimized packing of active site residues and ligand interactions within the

substrate binding pocket based on the geometry specified in the input catalytic array. To sample efficiently, FastDesign could only exchange amino acids within the active site and was constrained at each position to residue identities that were assigned high probabilities by ProteinMPNN (*48*). The generated sequences were predicted using ESMFold (*49*) and the highest scoring structures were fed into the next refinement cycle.

Following backbone refinement, an additional script utilized the CoupledMoves protocol (*50*, *51*) to refine ligand interactions to the binding pocket. ProteinMPNN was used with its soluble weights (*52*) to redesign the remaining protein backbone, generating the final enzyme sequences. In the final evaluation step, these sequences were predicted with AlphaFold2 and ranked according to a set of metrics (see design SI). We combined the scripts for generating artificial motif libraries, backbone diffusion and refinement, and CoupledMoves into a single method for semi-automated design of enzymes, called Riff-Diff (**R**otamer **i**nverted **f**ragment **f**inder – **Diff**usion, Figure 1B). Riff-Diff requires only minimal user input to generate sequences that scaffold a given catalytic array.

With the ability to accurately place arbitrary catalytic arrays within the binding pockets of *de novo* scaffolds, we investigated how effectively this approach produces enzymes experimentally. For this, we focused on a previously highly evolved catalytic tetrad of an artificial retro-aldolase and used it as our model system (PDB: 5AN7). This particular active site emerged after extensive rounds of laboratory evolution (*29*, *53*, *31*) and catalyzes the cleavage of R-methodol to 6-NMA and acetone at a rate close to natural class-1 aldolases. Mechanistic and mutational studies, as well as multiple computational analyses have illuminated the contributions of individual amino acids in this active site to its catalytic potency (*31*, *54–56*). This renders the tetrad an ideal candidate to evaluate the efficacy of 'transplanting' highly potent catalytic arrays to *de novo* scaffolds.

Riff-Diff reliably scaffolded backbones and precisely positioned the model tetrad (Figure 1E). Our newly designed sequences were predicted to fold into the designed structure, with average pLDDTs of 93 by AlphaFold2 and average backbone Cα-RMSDs of 0.57 Å (AF2). In addition to highly confident predictions, the artificial enzyme sequences exhibited favorable free folding energies (*57*) and low surface aggregation propensities (*58*) when evaluated with Rosetta FastRelax (*59*) (Figure 1C,D). Importantly, catalytic arrays of a large fraction of the designed enzymes exhibited predicted side chain RMSDs below one Angstrom compared to the input geometry. Design metrics were summarized in Figure 1D.

Scaffolding of a catalytic tetrad yielded highly proficient enzymes

Sequence-structure pairs designed by the Riff-Diff pipeline generally passed structural quality filters. We therefore predominantly ranked them by the predicted side chain RMSDs of the catalytic input geometry and by the interactions of the ligand with the binding pocket. Additionally, we excluded designs for which the substrate binding pocket appeared inaccessible to the substrate, because it was buried too deep inside the protein and occluded by bulky side chains. The final selection consisted of 36 sequences, which were derived from 12 unique backbones generated from 8 unique artificial motifs. Summary statistics describing *in silico* metrics of the selected sequences are included in Supplementary Figure S1. We ordered the corresponding 36 synthetic genes of which 35 could be cloned. In an initial 96 well-plate screen, all of the designs expressed soluble with the majority of variants exhibiting measurable catalytic activity, as followed via progression curves of product release. However, detailed Michaelis-Menten kinetics revealed that the activity varied drastically between variants with some exhibiting rate accelerations exceeding five-millionfold over the uncatalyzed reaction. To the best of our knowledge, this is higher than any previously reported *de novo* enzymes without cofactor designed zero-shot. An initial correlational analysis of the variants revealed that classic protein design metrics (e.g. structure-prediction confidence, catalytic side chain RMSD, ligand binding energy) failed to predict activity. To investigate the origins of the highly divergent catalytic rates between variants, we conducted further biochemical characterization of all variants from large scale expressions and recorded thermodynamic data.

Similar to the deep-well plate expression, batch fermentation yielded soluble protein for all designs and purification showed a monodisperse peak corresponding to the monomeric species on a gel filtration column (Figure 2A). The designs were produced in high yields (Figure 2C), with intact mass spectrometry analysis validating the identity of the purified proteins and CD spectroscopy confirming their helical structure and high thermal stability in solution (Figure 2B, Supplementary Figure S2). Furthermore, SAXS measurements confirmed the overall structure of our enzymes in solution. Dimensionless Kratky plots substantiate their high degree of overall foldedness as well as their globular shape (Supplementary Figure S3). 9 of the 35 experimentally tested constructs display a radius of gyration calculated from the Guinier region of the scattering profiles that is within 10% of that of the AF2-predicted structure. In 32 constructs, the difference is below 20% (Supplementary Table S1). A fit of scattering profiles calculated from the design models with FoXS (*60*, *61*) on the experimental scattering data reveals chi² values between 1.1 and 39.5, with the majority of samples (25 of 35) displaying

values below 2.6 (Figure 2C). These values are well below the chi² obtained from a fit of the scattering profile calculated from the experimental crystal structure of RAD32 (C_α RMSD of 1.21 Å to design model, chi² of 3.2), corroborating the agreement between design models and proteins in solution.

To further investigate the designs' thermodynamics, we measured resistance against chemical denaturation by guanidine hydrochloride (GdnHCl). 20 of 33 enzymes displayed cooperative unfolding with mid-points of denaturation ranging from 2.5 to > 6.5 M GdnHCl (Figure 2C, Supplementary Figure S4), indicating a wide range of thermodynamic stability. Thermodynamic stability is an important quality of enzymes used in industrial settings where they must endure harsh reaction conditions and the presence of organic cosolvents. To this end, we investigated whether computational metrics could predict the measured denaturation midpoints of our enzymes. Indeed, we found statistically significant correlations for prediction confidence by AlphaFold2 (pLDDT), free folding energy of the Rosetta Energy Function, core atomic density, and Rosetta's surface aggregation propensity (SAP) score (Supplementary Figure S5). Using these metrics, we built a simple linear regression model that confidently predicts chemical denaturation midpoints with a Pearson R of 0.9 (Figure 2D).

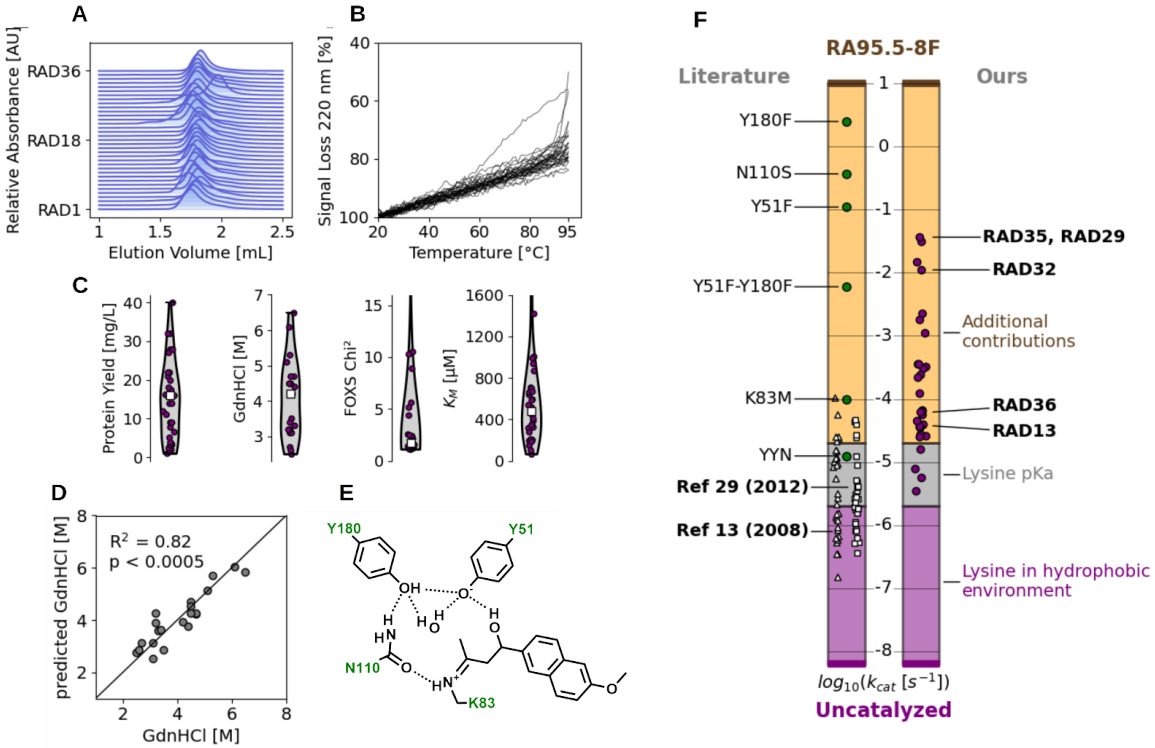

Figure 2: Activity of designed retro-aldol enzymes far exceeded those of previous zero-shot designs. A. All retro-aldolases eluted at the volume corresponding to the monomeric peak from the size exclusion chromatography column. SEC traces were normalized and stacked. **B.** CD melting curves confirmed high thermodynamic stability. With the exception of RAD23, only negligible loss in signal intensity at 220 nm can be observed up to 95°C. **C.** Yield, chemical denaturation resistance and agreement between experimental and calculated SAXS profiles. Designs were recombinantly produced in *Escherichia coli* with reasonable to good yields, ranging from 1 to 40 mg protein per liter of culture medium. Chemical denaturation midpoints for 20 of 35 constructs range from 2.5 to >6 M of guanidinium hydrochloride according to CD experiments. Fits of experimental SAXS data and scattering profiles calculated from design models show good agreement, with 25 of 35 designs displaying chi² values below 2.6. RAD5 (chi² of 39.5) was omitted from the plot. Affinity of the enzymes to the substrate ranged between 100 μM up to > 1500 μM. **D.** A linear regression model based on AlphaFold2 prediction confidence, aggregation propensity, and Rosetta score terms, confidently predicts chemical denaturation midpoints. **E.** Representation of the model tetrad, the schiff-base intermediate state of the reaction. **F.** The high activity of the designed enzymes indicates that the catalytic tetrad was functional. For most enzymes, rate accelerations exceed catalytic effects provided by the lysine in the hydrophobic pocket alone. On average, the designed enzymes were orders of magnitude faster than zero-shot designed retro-aldolases of previous design campaigns.

High activity originated from catalytic tetrad

The chosen active site from the evolved *de novo* retro aldolase RA95.5-8f cleaves methodol at a maximum velocity (k_{cat}) of 10 s^{-1}, which is more than 10^9-fold faster than the uncatalyzed reaction, occurring at 6.5 * 10^{-9} s^{-1} (*29*, *31*). We hypothesize that the rate resulted from a combination of functional coupling of the individual residues of this catalytic tetrad and conformational effects exerted by the surrounding active site environment (*54–56*). Given that our starting motifs were derived from this catalytic tetrad only and not its surrounding residues, the maximum expected k_{cat} of our designs should be lower, but approaching the reported rate. We determined Michaelis-Menten constants for 31 of the 35 produced retro-aldolases (Supplementary Figure S6). The remaining four designs did not display high enough activities to record detailed kinetics. We identified the most catalytically proficient variants to be RAD35 and RAD29. They catalyzed the cleavage of racemic methodol with a k_{cat} of 3.6 * 10^{-2} s^{-1} and 3.1 * 10^{-2} s^{-1}, a roughly 5 * 10^6-fold rate acceleration over the uncatalyzed reaction. Thus, our method successfully produced enzymes orders of magnitude faster than previous zero-shot designed retro aldolases, even surpassing the activity of the evolved catalytic antibody 38C2 (k_{cat} 1.1 * 10^{-2} s^{-1}) (*53*). Of particular note is the high affinity of RAD29 for (rac)-methodol. Having a K_M of ~100 μM, its catalytic efficiency of 290 M^{-1} s^{-1} approaches that of the extensively evolved RA95.5-5 (320 M^{-1} s^{-1}) (*53*).

Analysis of rate enhancements shows high activity likely originated from catalytic tetrad

Success rates of enzyme design campaigns like ours could greatly improve with the ability to predict enzyme activity computationally. This ability depends on a deep understanding of the origins of their rate accelerations. An analysis of the first computationally designed retro-aldolases demonstrated that lysines in hydrophobic pockets can achieve accelerations over the uncatalyzed reaction of up to 5 * 10^2. Additional 10-fold rate increases can be achieved by lowering the lysine's pKa (*30*). Evidence of catalysis by lysines in the hydrophobic environment of micelles provided further support for this finding (*62*), as well as the activity of a Y51F, Y180F, N110S variant of the evolved retro-aldolase RA95.5-8F, which falls into this range (Figure 3). Fourteen (45%) of our designed enzymes exhibited activities k_{cat}/k_{uncat} below 10^4-fold (Supplementary Table S2), suggesting that catalysis in those variants was mostly conferred by the lysine and not the surrounding residues of the tetrad. The essential role of the active site

lysine for catalysis in our designs was further corroborated by vastly diminished activity upon mutation to alanine (Supplementary Figure S7).

Beyond the acceleration conferred by the lysine schiff-base as an electron sink, C-C bond cleavage appears to be rate-limiting for the retro-aldol reaction (*54*, *56*). Seven of our designed enzymes accelerated methodol cleavage with k_{cat}/k_{uncat} over 10^5-fold, strongly indicating participation of the designed tetrad in the reaction (Figure 2A). The preference of RAD29 to form R-methodol in the forward reaction (aldol addition) further supports this conclusion (Supplementary Figure S8). Our results indicate that for seven (20%) of 35 enzymes, the scaffolded catalytic tetrad was at least partially functional, confirming the viability of our approach. Nevertheless, the similarity of our enzymes' computational metrics, in contrast to their clear difference in activity, was striking. Our findings underline the need for high resolution structural information to validate computational predictions, despite their recent gains in accuracy.

Crystal structures reveal unexpected relationship between structure and function

We set up crystallization screens and were able to obtain experimental structures of the four variants RAD13 (k_{cat} 3.8 * 10^{-5} s^{-1}), RAD17 (k_{cat} n.d.), RAD32 (k_{cat} 1.1 * 10^{-2} s^{-1}) and RAD36 (k_{cat} 6.2 * 10^{-5} s^{-1}). The variant's divergent catalytic rates are ideal to investigate how much of the variation can be explained by side chain packing in the active site. The crystal structures are in exceptional agreement with the design models, having an average overall backbone C_α RMSD of 0.89 Å (Figure 3A). The catalytic residues of RAD13 and RAD17 adopted the correct conformation with heavy atom side chain RMSDs of 0.42 Å and 1.09 Å to the design model and 0.89 Å and 1.2 Å to the catalytic geometry (Figure 3B). The near-atomic precision with which the tetrad was scaffolded in both variants is remarkable. However, despite the accurate side chain placement, the reaction rates of RAD13 and RAD17 fall far below the expected maximum rate. This indicates that precise positioning of catalytic side chains is a necessary design objective, but by itself not sufficient for the design of highly efficient biocatalysts.

In the active site of RAD36 we observed density for three distinct conformations of the catalytic lysine. One of the three lysine conformations points away from the active site, which explains its low catalytic rate. In contrast, in the active site of RAD32, the most active variant for which an experimental structure could be obtained, a tyrosine of the catalytic tetrad adopts a catalytically unproductive conformation entirely. This tyrosine is supposed to act as a catalytic base,

promoting C-C bond cleavage. However, Tyr99, another tyrosine in the RAD32 active site, is placed at a distance, which enables it to recover this interaction by having its hydroxyl group close to the position of the methodol β-alcohol in the model (Figure 3B).

In previous enzyme design campaigns, crystal structures have consistently shown that precise positioning of catalytic side chains was crucial for enzymatic activity (*23*, *63*, *64*). In contrast, our crystal structures revealed a discrepancy between precise positioning and activity. A possible explanation for this discrepancy is the exceptional accuracy of designed side chain conformations, which do not differ as much as in previous studies. This enables analysis of activity with sub-angstrom accuracy for the first time, revealing this accuracy-discrepancy. Another explanation is that the crystals were measured at cryo-temperatures, which is known to omit conformations only sampled at higher temperatures. Recent studies showed how side chain conformations present at cryo-temperatures correspond to conformations at a local thermodynamic minimum. As such, they do not portray the conformational dynamics underlying catalysis under reaction conditions (*65*–*67*). Alternatively, the precise side chain positioning in the active sites of RAD13 and RAD17 could favorably stabilize the schiff-base intermediate of the model geometry over its subsequent transition state, greatly increasing the reaction free energy barrier. Based on our crystal structures, high-resolution computational methods could disentangle these alternatives.

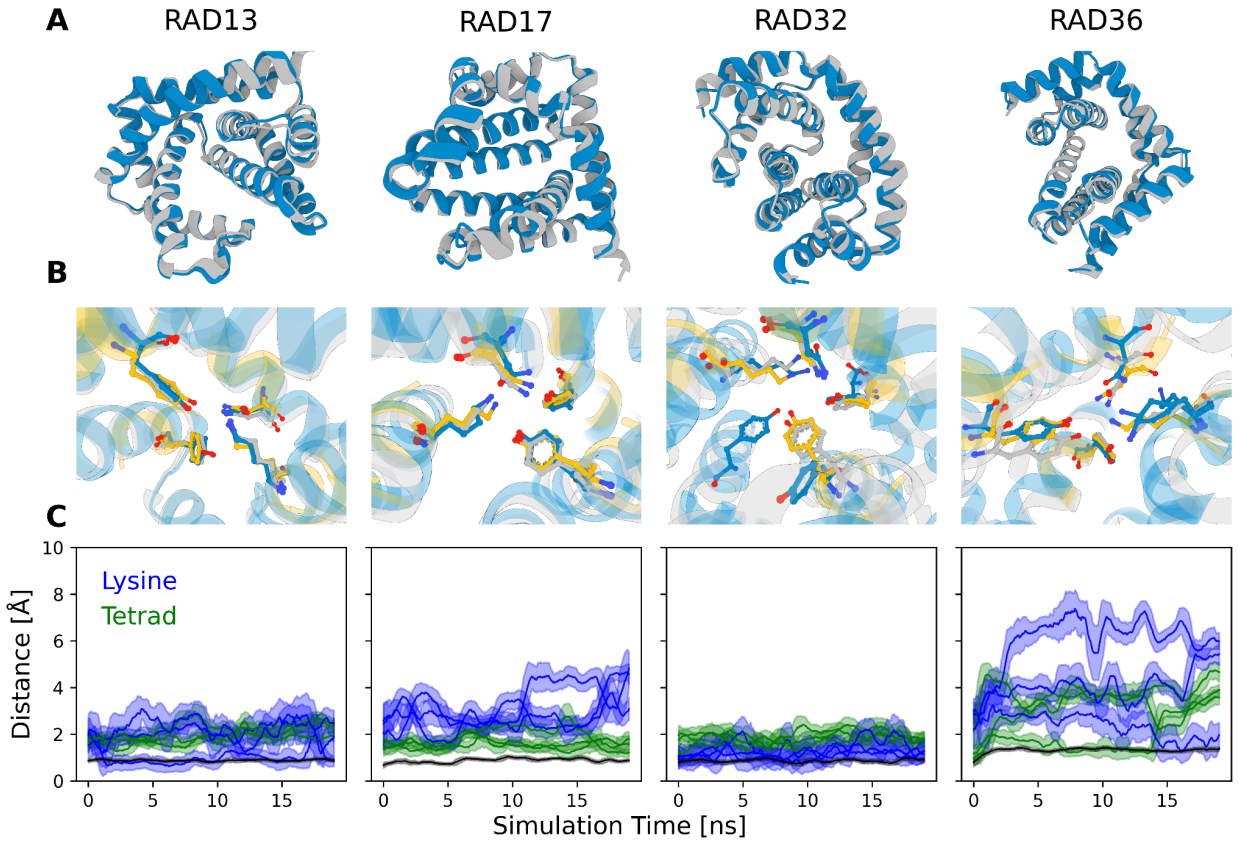

Figure 3. Crystal structures reveal exceptional accuracy of the scaffolded catalytic tetrad. A. The backbones of the design models (gray) are almost identical to the experimentally obtained crystal structures (blue), displaying overall C_α RMSDS below 1.2 Å. PDB IDs: 9GBT, 9FW5, 9FW7B, 9FWA. **B.** Active site residues in the crystal structures (blue) agree well with the design models (gray) and the theozyme (yellow). In the crystal structure of RAD32, the intended position of the tyrosine hydroxyl group is assumed by another tyrosine residue that was not part of the design model. Multiple conformations of the catalytic lysine residue are visible in the crystal structure of iRAD36. The conformer with highest occupancy is pointing away from the active site. **C.** MD simulations of the design models reveal varying degrees of preorganization of both the lysine (blue) and the tyrosine and asparagine residues of the tetrad (green) over simulation triplicates. The black traces correspond to the tetrad α-carbons.

Impact of preorganization on activity remains elusive

Conformational ensembles in enzyme active sites obtained from molecular dynamics simulations have frequently been used to explain experimental observations about activity. In theory, this conformational information could also be used predictively to screen enzyme variants before wet-lab experiments. For example, simulations revealed how the presence of catalytically unproductive conformational populations could differentiate active from inactive

kemp-eliminase designs (*63*). Further, an increase in preorganization correlated with improved catalysis along the evolutionary trajectory of designed retro-aldolase RA95 variants ultimately resulting in RA95.5-8F, the most efficient artificial retro-aldolase to date (*31*). We computed 20 ns triplicate molecular dynamics simulations for all 36 of our enzymes to evaluate how well preorganization around the theoretical geometry could predict activity. Triplicates of RA95.5-8F served as a positive control. Consistent with the literature, its catalytic tetrad was highly preorganized. In all replicates, the tetrad adopted a catalytically competent conformation, with the exception of Tyr180 (Supplementary Figure S9) which had an increased distance from its position in the crystal structure. We found that several of our variants exhibited a well preorganized active site geometry over the course of the trajectories. Yet, the degree of preorganization of the catalytic tetrad could not significantly differentiate highly active variants from variants with low activity. Many of the highly active variants even exhibited pronounced conformational flexibility of the catalytic lysine (Supplementary Figure S10). In contrast to the precise lysine conformations observed in the RAD13 and RAD17 crystal structures, the conformational populations sampled during the molecular dynamics trajectories were often far from the catalytically competent conformation. This conformational flexibility of the catalytic lysines explains the reduced activity of RAD13 and RAD17, which is not obvious at all from interpreting the crystal structures alone. Given the high variance between individual replicates and the inability to differentiate active from inactive designs, the significance of the information obtained from the molecular dynamics simulations remains speculative.

Conclusions and Outlook

Enzyme design and engineering in its current state relies on high-throughput screening methods to produce viable enzymes. While highly valuable, this method contradicts an original goal of the field, which was to produce catalysts for reactions that high-throughput screens cannot optimize. Our study constitutes a significant step forward in this direction. Using the semi-automated Riff-Diff method, we demonstrated that scaffolding of catalytic arrays yields highly active enzymes while screening as little as 35 enzyme sequences. An essential component for the high efficiency of our enzymes was to use a highly optimized catalytic center as a starting point. For novel reactions, building potent centers from scratch will require a robust model of how electrostatic preorganization, conveyed by individual functional groups, confers catalytic potency. In addition, static models of catalytic centers seem not to be sufficient by themselves as a design objective. This is highlighted by the contrast between the atomic accuracy of the catalytic center in RAD13 and its low catalytic activity. However, the ability to scaffold catalytic

centers in *de novo* proteins with high accuracy now allows the type of experiments that can elucidate such structure- and ensemble-function relationships, helping to build models of catalytic potency.

Next to potent catalytic centers, prediction of catalytic activity remains a bottleneck for zero-shot enzyme design. Protein language models represent a new frontier for this task but their poor generalizing abilities will limit their effectiveness for novel chemical reactions, especially with *de novo* enzymes. Physics-based methods like Quantum Mechanics / Molecular Mechanics (QM/MM) appear more suited to handle generalizability but they are computationally costly and require highly accurate structural data as inputs. This severely limits their predictive utility. However, conformational ensembles, obtained from molecular dynamics simulations, or otherwise (*68*, *69*), offer an alternative route for physics-based activity prediction. The simulations conducted in this study disclosed significant differences in the conformational populations of the catalytic tetrad between individual variants. Yet, with high variance between individual replicates of the trajectories, the conformational ensembles could not significantly discriminate between high- and low-activity variants. Furthermore, the replacement of a catalytic tyrosine at an alternative position in the crystal structure of RAD32 demonstrated how generic catalytic contributions within the active site must be considered in predictive models of activity. Thus, development of computationally cheap estimates for conformational populations, and generic models of catalytic contributions are critical for future improvements in this field.

Building models of generic catalytic contributions will require additional datasets like the one presented in this study. Thorough biochemical characterization allowed us to identify 32 enzymes that are folded correctly while exhibiting vastly divergent catalytic rates and substrate affinities. Analysis of the catalytic efficiency of the enzymes are supplemented with rich dynamic data obtained from molecular dynamics simulations and crystal structures for four of the 32 variants. The resulting dataset is an ideal test set to benchmark models of catalytic activity. For general models of reactivity, datasets like this will be needed for a diverse set of chemical transformations.

The ability to design enzymes like ours could foster the creation of blind tests for the prediction of enzymatic activity in artificial enzymes. Blind tests like this (CASP) give a clear picture of the state of the field and draw significant attention of the community to the relevant problems. Much like CASP fostered the development of now abundant methods for protein folding and design, a

contest for the blind prediction of enzyme activity could drive the development of methods for functional zero-shot design of enzymes.

Finally, predictions of enzyme activity have played a pivotal role in biocatalysis campaigns to create superior catalysts with tailor-made properties for the generation of new pharmaceuticals, and complex fine chemicals. Yet, very few of these studies started from scratch and all of the ones that did, relied on directed evolution to enhance activities in the designed biocatalysts. Here we established a method which shows great promise in transforming enzyme design from a field in which detailed domain knowledge is essential into an approach that the broader biotechnological community can apply.

ACKNOWLEDGEMENTS

We thank A. Winkler and P. Macheroux for discussions about (native) mass spectrometry and enzyme kinetics. Additionally, 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. M.B. and M.T. were supported by the Austrian Science Fund (FWF) grant 10.55776/DOC130 and trained within the framework of the PhD program Biomolecular Structures and Interactions (BioMolStruct). W.E. was supported by the Austrian Science Fund (FWF) grant 10.55776/P30826. A.T., M.C., D.S. and G.O. were supported by funding from the European Research Council through an ERC Starting Grant (HelixMold 802217) and a FETOPEN project (ARTIBLED, 863170). This research was funded in whole, or in part, by the Austrian Science Fund (FWF) [10.55776/P30826 to GO]. For open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Author contributions: Conceptualization: M.B., A.T., G.O.; Methodology: M.B., A.T., G.O.; Software: M.B., A.T.; Validation: M.B., A.T., W.E.; Formal analysis: M.B., A.T., S.K.; Investigation: M.B., A.T., M.C., S.K., M.T., D.S., A.B., W.E., S.Y.H., M.A., M.H., G.O.; Resources: M.A., M.H., G.O.; Data Curation: M.B., A.T., A.B., G.O.; Writing: M.B., A.T., G.O.; Visualization: M.B., A.T., S.K.; 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, command lines, Jupyter notebooks, and datasets will be made available upon peer-reviewed publication.

References

1. R. Buller, S. Lutz, R. J. Kazlauskas, R. Snajdrova, J. C. Moore, U. T. Bornscheuer, From nature to industry: Harnessing enzymes for biocatalysis. *Science* **382**, eadh8615 (2023).
2. A. H.-W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk, D. Baker, De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
3. Q. Wang, X. Liu, H. Zhang, H. Chu, C. Shi, L. Zhang, J. Bai, P. Liu, J. Li, X. Zhu, Y. Liu, Z. Chen, R. Huang, H. Chang, T. Liu, Z. Chang, J. Cheng, H. Jiang, Cytochrome P450 Enzyme Design by Constraining the Catalytic Pocket in a Diffusion Model. *Research* **7**, 0413 (2024).
4. I. Kalvet, M. Ortmayer, J. Zhao, R. Crawshaw, N. M. Ennist, C. Levy, A. Roy, A. P. Green, D. Baker, 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).
5. J. D. Schnettler, M. S. Wang, M. Gantz, H. A. Bunzel, C. Karas, F. Hollfelder, M. H. Hecht, Selection of a promiscuous minimalist cAMP phosphodiesterase from a library of de novo designed proteins. *Nat. Chem.* **16**, 1200–1208 (2024).
6. B. D. P. Mariz, S. Carvalho, I. L. Batalha, A. S. Pina, Artificial enzymes bringing together computational design and directed evolution. *Org. Biomol. Chem.* **19**, 1915–1925 (2021).
7. A. S. Klein, C. Zeymer, Design and engineering of artificial metalloproteins: from *de novo* metal coordination to catalysis. *Protein Eng. Des. Sel.* **34**, gzab003 (2021).
8. E. J. Hossack, F. J. Hardy, A. P. Green, Building Enzymes through Design and Evolution. *ACS Catal.* **13**, 12436–12444 (2023).
9. L. Pauling, Nature of Forces between Large Molecules of Biological Interest*. *Nature* **161**, 707–709 (1948).
10. A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci. Publ. Protein Soc.* **15**, 2785–2794 (2006).
11. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
12. J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **329**, 309–313 (2010).
13. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387–1391 (2008).
14. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).
15. S. Bjelic, L. G. Nivón, N. Çelebi-Ölçüm, G. Kiss, C. F. Rosewall, H. M. Lovick, E. L. Ingalls, J. L. Gallaher, J. Seetharaman, S. Lew, G. T. Montelione, J. F. Hunt, F. E. Michael, K. N. Houk, D. Baker, Computational Design of Enone-Binding Proteins with Catalytic Activity for the Morita–Baylis–Hillman Reaction. *ACS Chem. Biol.* **8**, 749–757 (2013).
16. H. Lechner, V. R. Emann, M. Breuning, B. Höcker, An Artificial Cofactor Catalyzing the Baylis‐Hillman Reaction with Designed Streptavidin as Protein Host**. *ChemBioChem* **22**, 1573–1577 (2021).
17. D. W. Watkins, J. M. X. Jenkins, K. J. Grayson, N. Wood, J. W. Steventon, K. K. Le Vay, M. I. Goodwin, A. S. Mullen, H. J. Bailey, M. P. Crump, F. MacMillan, A. J. Mulholland, G.

Cameron, R. B. Sessions, S. Mann, J. L. R. Anderson, Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. *Nat. Commun.* **8**, 358 (2017).
18. M. S. Wang, M. H. Hecht, A Completely *De Novo* ATPase from Combinatorial Protein Design. *J. Am. Chem. Soc.* **142**, 15230–15234 (2020).
19. A. E. Donnelly, G. S. Murphy, K. M. Digianantonio, M. H. Hecht, A de novo enzyme catalyzes a life-sustaining reaction in Escherichia coli. *Nat. Chem. Biol.* **14**, 253–255 (2018).
20. X. Chen, B. M. Discher, D. L. Pilloud, B. R. Gibney, C. C. Moser, P. L. Dutton, De Novo Design of a Cytochrome *b* Maquette for Electron Transfer and Coupled Reactions on Electrodes. *J. Phys. Chem. B* **106**, 617–624 (2002).
21. J. Kaplan, W. F. DeGrado, *De novo* design of catalytic proteins. *Proc. Natl. Acad. Sci.* **101**, 11566–11570 (2004).
22. D. N. Bolon, S. L. Mayo, Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci.* **98**, 14274–14279 (2001).
23. S. Rajagopalan, C. Wang, K. Yu, A. P. Kuzin, F. Richter, S. Lew, A. E. Miklos, M. L. Matthews, J. Seetharaman, M. Su, J. F. Hunt, B. F. Cravatt, D. Baker, Design of activated serine–containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10**, 386–391 (2014).
24. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
25. J. Z. Ruscio, J. E. Kohn, K. A. Ball, T. Head-Gordon, The Influence of Protein Dynamics on the Success of Computational Enzyme Design. *J. Am. Chem. Soc.* **131**, 14111–14115 (2009).
26. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
27. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
28. M. Baek, I. Anishchenko, I. R. Humphreys, Q. Cong, D. Baker, F. DiMaio, Efficient and accurate prediction of protein structure using RoseTTAFold2. [Preprint] (2023). https://doi.org/10.1101/2023.05.24.542179.
29. E. A. Althoff, L. Wang, L. Jiang, L. Giger, J. K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert, D. Baker, Robust design and optimization of retroaldol enzymes. *Protein Sci.* **21**, 717–726 (2012).
30. J. K. Lassila, D. Baker, D. Herschlag, Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci.* **107**, 4937–4942 (2010).
31. R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
32. A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524–8532 (2017).
33. H. A. Bunzel, J. L. R. Anderson, D. Hilvert, V. L. Arcus, M. W. Van Der Kamp, A. J.

Mulholland, Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat. Chem.* **13**, 1017–1022 (2021).
34. H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci.* **109**, 3790–3795 (2012).
35. G. Kiss, D. Röthlisberger, D. Baker, K. N. Houk, Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).
36. R. Lipsh-Sokolik, O. Khersonsky, S. P. Schröder, C. De Boer, S.-Y. Hoch, G. J. Davies, H. S. Overkleeft, S. J. Fleishman, Combinatorial assembly and design of enzymes. *Science* **379**, 195–201 (2023).
37. Y. Kipnis, A. O. Chaib, A. A. Vorobieva, G. Cai, G. Reggiano, B. Basanta, E. Kumar, P. R. E. Mittl, D. Hilvert, D. Baker, Design and optimization of enzymatic activity in a de novo β-barrel scaffold. *Protein Sci.* **31**, e4405 (2022).
38. S. Osuna, G. Jiménez-Osés, E. L. Noey, K. N. Houk, Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes. *Acc. Chem. Res.* **48**, 1080–1089 (2015).
39. M. V. Shapovalov, R. L. Dunbrack, A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **19**, 844–858 (2011).
40. J. B. Ingraham, M. Baranov, Z. Costello, K. W. Barber, W. Wang, A. Ismail, V. Frappier, D. M. Lord, C. Ng-Thow-Hing, E. R. Van Vlack, S. Tie, V. Xue, S. C. Cowles, A. Leung, J. V. Rodrigues, C. L. Morales-Perez, A. M. Ayoub, R. Green, K. Puentes, F. Oplinger, N. V. Panwar, F. Obermeyer, A. R. Root, A. L. Beam, F. J. Poelwijk, G. Grigoryan, Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023).
41. G. Huguet, J. Vuckovic, K. Fatras, E. Thibodeau-Laufer, P. Lemos, R. Islam, C.-H. Liu, J. Rector-Brooks, T. Akhound-Sadegh, M. Bronstein, A. Tong, A. J. Bose, Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation. arXiv [Preprint] (2024). https://doi.org/10.48550/ARXIV.2405.20313.
42. J. Wang, S. Lisanza, D. Juergens, D. Tischer, J. L. Watson, K. M. Castro, R. Ragotte, A. Saragovi, L. F. Milles, M. Baek, I. Anishchenko, W. Yang, D. R. Hicks, M. Expòsit, T. Schlichthaerle, J.-H. Chun, J. Dauparas, N. Bennett, B. I. M. Wicky, A. Muenks, F. DiMaio, B. Correia, S. Ovchinnikov, D. Baker, Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
43. R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. DiMaio, D. Baker, Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).
44. R. Wang, X. Fang, Y. Lu, S. Wang, The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **47**, 2977–2980 (2004).
45. O. J. Goudy, A. Nallathambi, T. Kinjo, N. Randolph, B. Kuhlman, *In silico* evolution of protein binders with deep learning models for structure prediction and sequence design. [Preprint] (2023). https://doi.org/10.1101/2023.05.03.539278.
46. P. Bryant, A. Elofsson, EvoBind: *in silico* directed evolution of peptide binders with AlphaFold. [Preprint] (2022). https://doi.org/10.1101/2022.07.23.501214.
47. J. B. Maguire, H. K. Haddox, D. Strickland, S. F. Halabiya, B. Coventry, J. R. Griffin, S. V. S. R. K. Pulavarti, M. Cummins, D. F. Thieker, E. Klavins, T. Szyperski, F. DiMaio, D. Baker, B. Kuhlman, Perturbing the energy landscape for improved packing during computational protein design. *Proteins Struct. Funct. Bioinforma.* **89**, 436–449 (2021).
48. J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky,

A. Courbet, R. J. De Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
49. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
50. N. Ollikainen, R. M. De Jong, T. Kortemme, Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. *PLOS Comput. Biol.* **11**, e1004335 (2015).
51. C. A. Smith, T. Kortemme, Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* **380**, 742–756 (2008).
52. C. A. Goverde, M. Pacesa, N. Goldbach, L. J. Dornfeld, P. E. M. Balbi, S. Georgeon, S. Rosset, S. Kapoor, J. Choudhury, J. Dauparas, C. Schellhaas, S. Kozlov, D. Baker, S. Ovchinnikov, A. J. Vecchio, B. E. Correia, Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
53. L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
54. C. Zeymer, R. Zschoche, D. Hilvert, Optimization of Enzyme Mechanism along the Evolutionary Trajectory of a Computationally Designed (Retro-)Aldolase. *J. Am. Chem. Soc.* **139**, 12541–12549 (2017).
55. A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524–8532 (2017).
56. D. De Raffele, S. Martí, V. Moliner, QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catal.* **9**, 2482–2492 (2019).
57. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
58. T. M. Lauer, N. J. Agrawal, N. Chennamsetty, K. Egodage, B. Helk, B. L. Trout, Developability Index: A Rapid In Silico Tool for the Screening of Antibody Aggregation Propensity. *J. Pharm. Sci.* **101**, 102–115 (2012).
59. L. G. Nivón, R. Moretti, D. Baker, A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS ONE* **8**, e59004 (2013).
60. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
61. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J.* **105**, 962–974 (2013).
62. J. Schmidt, C. Ehasz, M. Epperson, K. Klas, J. Wyatt, M. Hennig, M. Forconi, The effect of the hydrophobic environment on the retro-aldol reaction: comparison to a computationally-designed enzyme. *Org. Biomol. Chem.* **11**, 8419 (2013).
63. H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci.* **109**, 3790–3795 (2012).
64. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S.

Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
65. G. Wei, W. Xi, R. Nussinov, B. Ma, Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.* **116**, 6516–6551 (2016).
66. F. Yabukarski, T. Doukov, M. M. Pinney, J. T. Biel, J. S. Fraser, D. Herschlag, Ensemble-function relationships to dissect mechanisms of enzyme catalysis. *Sci. Adv.* **8**, eabn7738 (2022).
67. F. Yabukarski, J. T. Biel, M. M. Pinney, T. Doukov, A. S. Powers, J. S. Fraser, D. Herschlag, Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray–derived conformational ensembles. *Proc. Natl. Acad. Sci.* **117**, 33204–33215 (2020).
68. B. Jing, B. Berger, T. Jaakkola, AlphaFold Meets Flow Matching for Generating Protein Ensembles. arXiv [Preprint] (2024). https://doi.org/10.48550/ARXIV.2402.04845.
69. C. Duran, G. Casadevall, S. Osuna, Harnessing conformational dynamics in enzyme catalysis to achieve nature-like catalytic efficiencies: the shortest path map tool for computational enzyme redesign. *Faraday Discuss.*, 10.1039.D3FD00156C (2024).