Customizing the Structure of Minimal TIM Barrels to Craft Efficient *De Novo* Enzymes

Julian Beck,[1,†] Benjamin J. Smith,[2,3,†] Mark Kriegel,[1] Niayesh Zarifi,[2,3] Emily Freund,[1] Ahana G. Harsha,[2,3] Jan Hartmann,[1] Roberto A. Chica,[2,3,*] Birte Höcker[1,*]

[1] Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.

[2] Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5

[3] Center for Catalysis Research and Innovation, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5

[†] These authors contributed equally.

*** Corresponding authors**

Birte Höcker, E-mail: birte.hoecker@uni-bayreuth.de

Roberto A. Chica, E-mail: rchica@uottawa.ca

Keywords

De novo enzyme design, computational protein design, enzymes, TIM barrel, Kemp elimination

Abstract

The TIM barrel is the most prevalent fold in natural enzymes, supporting efficient catalysis of diverse chemical reactions. While *de novo* TIM barrels have been successfully designed, their minimalistic architecture lacks structural elements essential for substrate binding and catalysis. Here, we present CANVAS, a computational workflow that introduces a structural lid into a minimal *de novo* TIM barrel to anchor catalytic residues and form an active-site pocket for enzymatic function. Starting from two *de novo* TIM barrels, we designed nine variants with distinct lids to form active sites for the Kemp elimination. Four designs showed measurable activity, with the most active reaching a catalytic efficiency of 21,000 M^{-1} s^{-1} at its optimal pH. A co-crystal structure of this variant bound to a transition-state analogue confirmed the accuracy of the designed lid and active site. Using the X-ray structure of a lower-activity variant (19 M^{-1} s^{-1}), we applied ensemble-based design to optimize its active site, increasing catalytic efficiency by >1,600-fold to 32,000 M^{-1} s^{-1}. These results demonstrate that *de novo* TIM barrels can be endowed with substrate binding pockets supporting efficient catalytic function, establishing a platform for building enzymes on demand from minimal protein scaffolds.

The TIM barrel is one of the most versatile enzyme folds. It is present in six of the seven enzyme classes and supports catalysis at diffusion-limited rates[1–3]. For these reasons TIM barrels have frequently been used as scaffolds for enzyme design[4–7]. While idealized *de novo* TIM barrels have been successfully designed[8–11], none display enzymatic function due to their minimalistic architecture, which lacks key structural elements required for catalysis, such as cavities, pockets, and extended loops. Efforts to address these deficiencies have led to the introduction of secondary structural elements to create rudimentary pockets[12–14]. However, these pockets are too solvent exposed and undefined to support the microenvironments required for catalysis. *De novo* TIM barrels have also been fused to additional protein domains to create hybrid architectures with metal-binding sites positioned at domain interfaces[15]. While these constructs demonstrate that domain fusion can generate large cavities acting as reaction chambers for photoredox catalysis[16], they rely largely on the inherent reactivity of the bound metal and lack substrate-specific active sites characteristic of natural TIM barrel enzymes. Therefore, novel strategies are needed to craft custom active sites and endow *de novo* TIM barrels with true enzymatic function.

A computational approach for active-site scaffolding

Here, we introduce CANVAS (Customizing Amino-acid Networks for Virtual Active-site Scaffolding), a computational workflow for transforming minimal proteins such as *de novo* TIM barrels into functional enzymes (Figure 1a, Supplementary Figure 1). The process begins with a minimal TIM-barrel template and a theozyme, which is a computational model of an idealized enzyme active site with catalytic groups arranged to stabilize a reaction's transition state[17]. Using the protein design software Triad[18], a catalytic amino acid from the theozyme is placed onto the TIM-barrel template, and the transition state is built from its side chain to maintain the catalytic geometry[19]. A second catalytic residue is then built from the transition state, ensuring proper

60 geometry for catalytic interaction. The C_α of this second residue is thus positioned in the empty
61 space above the TIM barrel catalytic face. This allows it to serve as an anchor for constructing a
62 custom lid to form the active-site pocket, which is designed using the generative AI tool
63 RFdiffusion[20] followed by sequence optimization with ProteinMPNN[21] and structure prediction with
64 AlphaFold2[22]. The active site is then optimized with Triad to build a pocket complementary to the
65 transition state[6]. Designs are filtered using key enzyme design criteria[23], including solvent-
66 accessible surface area of the transition state, active-site preorganization, energy, and catalytic
67 contact geometry.
68

69 **Applying CANVAS to minimal TIM-barrel proteins**

70 We applied CANVAS to the Kemp elimination (Figure 1b), a model organic transformation
71 commonly used as a benchmark for *de novo* enzyme design[4,6,23]. The reaction's theozyme
72 features an Asp as catalytic base and an Asn or Gln as H-bond donor to stabilize the binding of
73 the substrate and properly orient it for reaction with the base. We used the crystal structures of
74 two *de novo* TIM barrels (PDB IDs: 7MCD[24] and 7OSV[9]) as templates. Since 7OSV has its N-
75 and C-termini on the catalytic face of the TIM barrel, we applied an *in silico* circular permutation
76 based on the inpainting method[25] to generate two modified structures termed NT6-CP1 and NT6-
77 CP2 that contain termini on the stability face (Supplementary Figure 2). Starting from these
78 structures, we used CANVAS to generate distinct lids for each template tailored to the Kemp
79 elimination (Figure 1c). The lid diffused onto 7MCD consists of a single 26-residue fragment
80 introduced in βα-loop 7 (Supplementary Figure 3). By contrast, the designed lids in NT6-CP1 and
81 NT6-CP2 comprise a major fragment of 33 or 39 residues introduced into βα-loops 2 or 6,
82 respectively, to scaffold the catalytic H-bond donor, as well as three or four elongated loops to
83 enhance interactions with the primary fragment. The resulting designs include variants 1–2
84 obtained from NT6-CP2, variants 3–5 from NT6-CP1 and variants 6–9 from 7MCD, each with a

85 unique active-site configuration (Supplementary Figure 4) and sequence (Supplementary Table
86 1), which yielded high AlphaFold2 pLDDT scores (Supplementary Figure 5).
87
88 **Several active Kemp eliminases identified**
89 Of the nine designed variants, those derived from NT6-CP1 and NT6-CP2 (variants 1–5)
90 were either non-expressible or poorly soluble, whereas 7MCD-derived constructs (variants 6–9)
91 were readily expressed and soluble. Of these initial designs, only variant 4 showed detectable
92 activity; however, detailed characterization was hampered by low protein yields and a tendency
93 to aggregate (Supplementary Figure 6, Supplementary Table 2). The NT6-CP1 template itself
94 also failed to express, in contrast to NT6-CP2 and 7MCD. To improve expression and solubility
95 of variants 1–5, we used LigandMPNN to redesign all non-active-site residues while preserving
96 the Triad-designed active site. These new sequences are henceforth termed KempTIMs (Figure
97 2a, Supplementary Figure 7a, Supplementary Table 1). This strategy, previously applied to
98 stabilize natural enzymes[26], restored expression or increased solubility in all but one variant. The
99 four soluble KempTIMs were purified at yields of 7–59 mg L^{-1} (Supplementary Table 2) as
100 monomeric species (Figure 2b, Supplementary Figure 7b, Supplementary Table 3) with mixed
101 α/β secondary structure characteristic of TIM barrels (Figure 2c, Supplementary Figure 7c). All
102 displayed melting temperatures above 65 °C, and several remained folded at temperatures
103 exceeding 90 °C (Figure 2d, Supplementary Figure 7d).
104 Kinetic analysis revealed activity for all expressed KempTIMs (Figure 3a, Supplementary
105 Figure 8, Supplementary Table 4). The most active design, KempTIM1, exhibited a catalytic
106 efficiency of 1,400 M^{-1} s^{-1} at pH 7 (Table 1). Mutation of the catalytic base abolished activity
107 (Figure 3b), whereas mutation of the H-bond donor decreased k_{cat} sixfold (Figure 3c, Table 1),
108 without disrupting the structure (Supplementary Figure 9), confirming the intended functional roles
109 of the designed residues. A pH-rate profile (Supplementary Figure 10, Supplementary Table 5)

110 revealed increasing activity at alkaline pH, with a single pK_a of approximately 9.0, consistent with
111 the catalytic base Asp residing in a hydrophobic active-site environment. At its optimal pH of 10,
112 k_{cat}/K_M increased to 21,000 M^{-1} s^{-1} (Figure 3d, Table 1). KempTIM1 displayed k_{cat} values of 4.2 s^{-1}
113 at pH 7.0 and 13.7 s^{-1} at pH 10.0, representing 10–14-fold increases compared to recently
114 reported first-round Kemp eliminases assayed under similar conditions[27,28]. Together, these
115 results show that CANVAS reliably produces functional *de novo* enzymes and that diverse lid
116 architectures with distinct catalytic motifs can support Kemp elimination. Notably, the top-
117 performing KempTIM surpasses the catalytic efficiencies of first-round *de novo* Kemp eliminases
118 produced through alternative computational design strategies[27,28].
119
120 **Confirming the lid and active-site features**
121 To assess structural accuracy of the CANVAS-designed lids and their associated active
122 sites, we solved crystal structures (Supplementary Table 6) of the most efficient design,
123 KempTIM1 (21,000 M^{-1} s^{-1}), and a lower-efficiency variant incorporating a different lid, KempTIM4
124 (19 M^{-1} s^{-1}). In both enzymes, the designed lids fold as intended, with the helix–turn–helix motif
125 and elongated loops closely matching the design models (lid backbone Cα RMSD of 0.9 Å for
126 KempTIM1 and 1.2 Å for KempTIM4; Figure 4a,c). In both cases, however, the lid is slightly
127 displaced relative to the TIM-barrel core, shifting the Cα position of the catalytic H-bond donor by
128 1.0 Å in KempTIM1 and 2.0 Å in KempTIM4 relative to the design models (Figure 4b,d). Because
129 extensive crystal contacts were observed in the lid region of KempTIM4, we probed its solution
130 structure using small-angle X-ray scattering (SAXS). Comparison of different structural models
131 with the SAXS data indicates that the lid adopts a conformation in solution similar to that observed
132 in the crystal structure (Supplementary Figure 11).
133 Despite crystallizing both proteins in the presence of the transition-state analogue 6-
134 nitrobenzotriazole (6NBT), we observed clear electron density for this ligand only in the active site

135 of KempTIM1 (Supplementary Figure 12), indicating weak binding in KempTIM4 under the
136 crystallization conditions. In the KempTIM1 co-crystal structure (Figure 4b), active-site residues
137 adopt conformations within 0.8 Å RMSD of the design model. The designed catalytic base D24
138 adopts a rotamer similar to that in the design model, enabling formation of a hydrogen bond with
139 the bound transition-state analogue despite the ligand being slightly rotated relative to its
140 designed orientation. This productive conformation of D24 is stabilized by an ordered water
141 molecule (Supplementary Figure 12, Supplementary Figure 13). By contrast, the H-bond donor
142 N131 adopts an alternative rotamer yet still forms a hydrogen bond with 6NBT, consistent with
143 the observed sixfold decrease in catalytic efficiency when this residue is mutated to alanine
144 (Figure 3c, Table 1). Notably, in the apo structure of KempTIM1 (Supplementary Figure 14), N131
145 adopts a rotamer that more closely matches the designed conformation and exhibits a Cα
146 displacement of 1.1 Å relative to the co-crystal structure, placing it near its location in the design
147 model. This backbone shift indicates a lid rearrangement upon substrate binding, a behavior
148 commonly observed in natural enzymes. This conformational adjustment is accompanied by a
149 rotameric change in the neighbouring residue R130, whose side chain reorients toward the
150 binding pocket in the co-crystal structure and forms a water-mediated hydrogen bond with the
151 transition-state analogue (Supplementary Figure 13).

152 The KempTIM1 co-crystal structure further shows a deviation of the transition-state
153 analogue nitro group. Notably, the catalytic residues form hydrogen bonds that better
154 accommodate a tautomeric form of 6NBT (Supplementary Figure 12, Supplementary Figure 15)
155 corresponding to the 6-nitrobenzisoxazole substrate rather than the 5-nitro regioisomer targeted
156 in the design. To determine whether this binding preference translates into altered regioselectivity,
157 we measured kinetic parameters using 6-nitrobenzisoxazole as substrate. KempTIM1 exhibits an
158 approximately twofold higher k_{cat}/K_M for 6-nitrobenzisoxazole compared to the 5-nitro isomer

159 (Supplementary Figure 16). These findings are consistent with the relatively large active-site
160 pocket of KempTIM1, which would accommodate both regioisomers.

161 In contrast to KempTIM1, most binding pocket residues in KempTIM4 adopt conformations
162 that differ from those intended to stabilize the transition state, indicating limited preorganization
163 of the active site (Figure 4d). While the catalytic base D153 adopts the designed rotamer, the H-
164 bond donor N46 adopts a conformation incompatible with transition-state stabilization. Consistent
165 with this structural observation, mutation of N46 to alanine results in only a minor decrease in
166 catalytic activity (Figure 3c, Table 1). Together, these structural features likely contribute to the
167 orders-of-magnitude lower catalytic efficiency of KempTIM4 relative to KempTIM1 and provide
168 clear targets for further optimization.

169

170 **Catalytic efficiency improved by ensemble-based design**

171 To correct active-site deficiencies observed in KempTIM4, we applied an ensemble-based
172 design[23,29] strategy (Supplementary Figure 17). This approach uses protein backbones generated
173 by crystallographic ensemble refinement[30] as design templates. It can reveal conformational
174 substates that better accommodate the transition state and its catalytic interactions, thereby
175 enabling the accurate construction of enhanced active sites[23].

176 During ensemble-based design, the identity of the catalytic base was fixed as Asp but its
177 position within the β-barrel was allowed to vary (Methods). The identity of the H-bond donor was
178 permitted to vary between Asn and Gln but could only be placed within the designed lid. Following
179 sequence design with Triad, we filtered the design models (Methods) and predicted 6NBT-bound
180 structures using the generative AI model Boltz-2[31] (Supplementary Figure 18). This filtering
181 pipeline yielded eight redesigned variants (KempTIM4a–h) comprising one of three distinct
182 catalytic motifs (Supplementary Table 1, Supplementary Table 4, Supplementary Figure 4), none
183 of which reproduced the motif found in the parent KempTIM4 (Table 1).

184 All second-round designs expressed solubly (Supplementary Table 2) and exhibited
185 increased catalytic efficiency relative to KempTIM4 at pH 7, with two variants displaying
186 improvements of at least two orders of magnitude (Supplementary Table 4, Supplementary Figure
187 8). The most active design, KempTIM4b, expressed as a stable monomeric TIM barrel (Figure 2)
188 and reached a k_{cat}/K_M of 3100 M^{-1} s^{-1} (Figure 3a), which is a >400-fold improvement thereby
189 surpassing the most active first-round design KempTIM1 (Table 1). Mutation of the catalytic base
190 of KempTIM4b abolished activity (Figure 3b), whereas substitution of the designed H-bond donor
191 with alanine reduced catalytic efficiency approximately fivefold (Figure 3c). This result contrasts
192 sharply with the parent KempTIM4, in which mutation of the H-bond donor had little effect on
193 activity. Neither mutation altered the overall protein fold (Supplementary Figure 9). Like
194 KempTIM1, KempTIM4b also showed enhanced activity at alkaline pH, with an apparent pK_a of
195 9.0 (Supplementary Figure 10, Supplementary Table 5). At pH 10, k_{cat} increased to 26 s^{-1} and
196 k_{cat}/K_M improved >1600-fold reaching 32,000 M^{-1} s^{-1} (Table 1, Figure 3d).
197 Notably, KempTIM4b reacted more efficiently with the intended target substrate 5-
198 nitrobenzisoxazole than with 6-nitrobenzisoxazole, exhibiting 5- or 8-fold higher k_{cat}/K_M at pH 7
199 and 10, respectively (Supplementary Table 4, Supplementary Figure 16). Together, the high
200 catalytic efficiency and substrate selectivity of KempTIM4b demonstrate that *de novo* TIM barrels
201 equipped with CANVAS-designed lids can host fully functional, selective active sites, and provide
202 a platform for further optimization toward even more powerful catalysts.
203
204 **Active-site dynamics distinguish high- and low-activity designs**
205 To probe the origins of activity differences among the designed enzymes, we performed
206 molecular dynamics simulations of all active KempTIMs bound to the transition-state analogue
207 6NBT, using their design models as starting structures. Across all designs, the lids were more
208 flexible than the minimal TIM-barrel core, but the helix anchoring the designed H-bond donor

209 remained comparatively rigid (Supplementary Figure 19). Notably, this helix was most rigid in the
210 highly active first-round design, KempTIM1, suggesting that increased helical rigidity may help
211 stabilize the substrate and maintain catalytically productive geometry.

212 To test this hypothesis, we evaluated the ability of designed active sites to stabilize the
213 transition-state analogue in a catalytically competent pose. For each trajectory snapshot, we
214 calculated both the distance and angle of the two designed catalytic contacts (Figure 5a). The
215 most active variants displayed narrow distributions centred on ideal values[32] (Supplementary
216 Figure 20), whereas lower-activity variants showed broadened distributions for one or both
217 contacts. The H-bond donor interaction was generally the least consistently formed, in agreement
218 with the observed lid displacement in the crystal structures (Figure 4) as well as the mutational
219 studies (Figure 3b,c).

220 To assess whether differences in the ability of active sites to stabilize the designed
221 catalytic contacts were related to the experimentally measured catalytic efficiencies, we
222 calculated a catalytic competency score[23], defined as the fraction of simulation snapshots in which
223 both catalytic contacts satisfied geometric criteria expected for efficient catalysis. Catalytic
224 competency scores correlated with measured k_{cat}/K_M values across the variant set (R = 0.76,
225 Figure 5b), with the two most active enzymes, KempTIM1 and KempTIM4b, exhibiting the highest
226 scores. These findings indicate that this metric may serve as a useful in-silico filter for prioritizing
227 future enzyme designs prior to experimental testing.

228

229 **Opportunities for CANVAS design**

230 Natural TIM barrels use lids to bind substrates and form active sites conducive to
231 catalysis[33]. By contrast, minimal *de novo* TIM barrels lack such lids and are inherently
232 functionless. Using our CANVAS approach, we demonstrate that custom lid construction can
233 transform these inert scaffolds into efficient enzymes. By designing various lids to simultaneously

234 anchor a catalytic residue and form a substrate binding pocket, we created a series of *de novo*
235 Kemp eliminases. One of these, KempTIM1, ranks among the most active *de novo* Kemp
236 eliminases whose active site was obtained from a single round of computational design. Notably,
237 its catalytic efficiency is 7-fold higher than that of the most active Kemp eliminase generated using
238 scaffolds assembled combinatorially from fragments of homologous TIM-barrel templates[27]. This
239 result demonstrates that minimal *de novo* TIM barrels equipped with custom-designed lids can
240 yield enzymes that outperform those generated using fragments derived from natural scaffolds.
241 Even variants with modest activity, such as KempTIM4, perform on par with many first-generation
242 enzymes produced by conventional theozyme-based methods[4,6,34], despite the added complexity
243 of engineering a tailored lid comprising multiple loops to create an active site cavity. Importantly,
244 CANVAS designs are evolvable. Using the crystal structure of KempTIM4 for ensemble-based
245 design[23], we increased catalytic efficiency by more than two orders of magnitude. Further
246 improvements to activity of KempTIMs might be achieved using experimental optimization, for
247 example NMR-guided directed evolution[35], which has previously uncovered distal mutations that
248 boost Kemp elimination rates in *de novo* enzymes to values approaching the diffusion limit[28].
249 CANVAS also addresses a fundamental limitation of *de novo* enzyme design strategies.
250 Unlike traditional approaches that retrofit catalytic groups into pre-existing folds[4,6], CANVAS
251 enables the modular customization of lids to satisfy the geometric and chemical requirements of
252 virtually any reaction. This design versatility becomes more critical as the number of catalytic
253 residues grows, a regime where fixed scaffolds make optimization increasingly difficult. While
254 recent generative-AI methods address this problem by building entirely new folds around catalytic
255 motifs extracted from the PDB or theozymes[36,37], yielding functional *de novo* enzymes, the
256 resulting proteins often bear little resemblance to natural enzymes. CANVAS offers a
257 complementary strategy: it retains a familiar, well-behaved scaffold while precisely tailoring only
258 the features essential for catalysis.

In the context of TIM barrels, this approach maintains a fold that is inherently versatile and widely used in nature across diverse chemistries. The β-barrel provides a stable framework for anchoring one or more catalytic residues, whereas the lid offers a tuneable element for controlling substrate binding and anchoring additional catalytic residues. Previous enzyme-design efforts have sought to harness this versatility by repurposing natural TIM barrels[4–6] or by recombining natural barrel fragments *in silico*[27], approaches that inherently restrict design space to existing lid architectures. By contrast, CANVAS decouples lid design from natural precedent, enabling the creation of bespoke lids tailored to specific catalytic demands. As a result, CANVAS transforms minimal, and even purely computationally generated, TIM barrels into genuine blank canvases for enzyme construction. We anticipate that this approach will unlock the full catalytic potential of designed TIM-barrel proteins across diverse chemical transformations. More broadly, the principles underlying CANVAS should extend beyond TIM barrels to other minimal protein architectures[38,39], where designed lids could be used to configure and stabilize one or more catalytic residues for new-to-nature catalysis.

Methods

Reagents and solutions. All experiments were conducted using analytical-grade chemicals, with solutions prepared using double-distilled water.

Circular permutation of DeNovoTIM6-SB with inpainting. Circular permutation of DeNovoTIM6-SB (PDB ID: 7OSV) was performed using the inpainting method described by Wang et al.[25] The first residue of either the first or third β-strand was selected as the new N-terminus, with the corresponding βα-loop deleted (Supplementary Figure 2). A new connection for the corresponding βα-loop was modeled by inpainting 5 to 10 residues. Unresolved residues in the

284 crystal structure were remodeled, keeping intact the positions and identities of the remaining
285 residues. For each new connection, 100 designs were modeled, followed by relaxation and
286 scoring using the Rosetta protein design software[40]. Structures of each design were predicted
287 using AlphaFold2[22] with model 4 weights. Designs with a pLDDT value above 90 and a Rosetta
288 score below 3 Rosetta energy units per residue were kept. One design from each connection was
289 selected for experimental characterization (Supplementary Table 1). The proteins were named
290 NT6-CP1 (N-terminus on the first β-strand) and NT6-CP2 (N-terminus on the third β-strand).
291
292 ***Computational enzyme design.*** All calculations were performed with the Triad protein design
293 software[18] (Protabit, Pasadena, CA, USA). Rotamer optimization was performed using a Monte
294 Carlo with simulated annealing search algorithm. Input structures were prepared for Triad
295 calculations via the *addH.py* application within Triad. The Kemp elimination transition state (TS)
296 structure was built using parameters from Privett *et al.*[6]. To provide sidechain conformations, a
297 backbone-independent conformer library (bda-bbind_1.0)[19] was used for theozyme placement,
298 and a backbone-independent rotamer library (bbind02.May.e2)[41] with expansions of ± 1 standard
299 deviation around χ_1 and χ_2 was used for active-site repacking. Energies were calculated using a
300 modified version of the Phoenix energy function[18,42] consisting of a Lennard-Jones 12–6 van der
301 Waals term from the Dreiding II force field[43] with atomic radii scaled by 0.9, a direction-dependent
302 hydrogen bond term with a well depth of 8.0 kcal mol^{-1} and an equilibrium donor–acceptor
303 distance of 2.8 Å,[44] and an electrostatic term modeled using Coulomb's law with a distance-
304 dependent dielectric of 10. An energy benefit of –100 kcal mol^{-1} was applied when TS-side-chain
305 interactions satisfied catalytic contact geometries (Supplementary Table 7), as described by
306 Lassila *et al.*[19]
307

308 ***Theozyme placement.*** AlphaFold2 predictions of NT6-CP1 and NT6-CP2, along with the 7MCD
309 crystal structure, were used as backbone templates for theozyme placement. Amino-acid
310 residues with a C_α–C_β vector oriented towards the β-barrel interior and located at the C-terminus
311 of each β-strand were selected as positions for introduction of the catalytic base (Asp).
312 Neighbouring residues were mutated to alanine to avoid steric clashes with the TS
313 (Supplementary Table 8). TS poses were built using the contact geometries listed in
314 Supplementary Table 9, and those where the isoxazolic oxygen pointed toward the catalytic face
315 of the TIM barrel and remained coplanar with the catalytic base were selected (Supplementary
316 Figure 1). In parallel, the catalytic H-bond donor (Asn or Gln) was placed on a Gly-Asn/Gln-Gly
317 tripeptide scaffold, and TS poses were built using H-bond donor side-chain contacts specified in
318 Supplementary Table 9. The resulting TS was superimposed with the TS from the catalytic base
319 calculation to create the full theozyme, positioning the catalytic base within the minimal TIM barrel
320 and the H-bond donor in the empty space above the catalytic face. The Gly residues from the
321 tripeptide were then deleted, preserving only the Asn/Gln residue, whose alpha carbon atom
322 served as an anchor for the designed lid.
323
324 ***Lid design.*** For all theozyme placements, peptide connections between the catalytic H-bond
325 donor residue and the minimal TIM barrel structures were designed using RFdiffusion[20]. For each
326 position of the H-bond donor alpha carbon, various βα-loops were selected as insertion points for
327 a primary peptide fragment, and multiple fragment lengths were sampled to connect to the
328 catalytic residue while preserving the TIM-barrel topology. Additional βα-loops were sometimes
329 elongated to improve interactions with the primary fragment (Supplementary Figure 3). For each
330 inserted primary fragment, 100–200 structures were sampled. Primary fragments lacking
331 secondary structure, causing steric clashes with the TS, or displacing the H-bond donor from its
332 theozyme position, were discarded (Supplementary Figure 1). The filtered fragments were

333 sequence-optimized using ProteinMPNN[21] using a temperature factor of 0.1, ensuring the minimal
334 TIM barrel template sequence remained unchanged. Cysteine and methionine residues were
335 excluded during sequence design. Structures of designed sequences were predicted using
336 AlphaFold2[22] or ColabFold v1.3.0[45] with all five model weights, and those displaying an average
337 pLDDT value > 90 and backbone RMSD < 3 Å relative to the RFdiffusion model were selected for
338 active-site repacking.
339
340 ***Active-site repacking.*** AlphaFold2 models generated above were used for a theozyme
341 placement step[19], as described previously, to verify that contacts between catalytic residues and
342 the TS could be formed on these scaffolds. Active-site repacking calculations were then
343 performed on the structures with this new theozyme placement. In these calculations, the TS was
344 translated by ± 0.4 Å along each Cartesian coordinate in 0.2-Å steps and rotated about all three
345 axes (origin at the TS geometric center) in 5° increments over a 10° range (clockwise and
346 counterclockwise). This resulted in a total combinatorial search space of 15,625 possible poses.
347 Residues near the catalytic amino acids and TS were designated as design positions. During
348 repacking, these positions were allowed to sample rotamers of various amino acids, favoring
349 hydrophobic and aromatic residues, while the identities of the catalytic residues remained fixed
350 (Supplementary Table 10).
351
352 ***Computational library design***. After repacking, the CLEARSS computational library design
353 algorithm[46] was used to generate a combinatorial library comprising the most favorable amino
354 acids predicted by Triad at each designed active-site position. In this method, libraries of a specific
355 size configuration are generated from a pre-scored list of sequences using the highest probability
356 set of amino acids at each position based on the sum of their Boltzmann weights. Using the list
357 of energy-ranked sequences from active-site repacking as input, libraries of 192 sequences were

358 generated. Rotamer configurations for each sequence in the library were optimized using the
359 *cleanSequences.py* application within Triad to find the lowest-energy conformation of each
360 sequence on its respective backbone, generating “cleaned” structures. To compare energies, the
361 energy difference between each “cleaned” structure and a corresponding all-Gly structure in the
362 absence of the TS was calculated. These energies are reported in Supplementary Table 11.
363 Structures were cleaned with and without the TS to evaluate preorganization, as described
364 previously[23].
365
366 ***Design filtering***. Energies of the TS-bound and unbound “cleaned” structures were calculated
367 using Triad. Dihedral angles of catalytic residues relative to the TS, as well as the solvent-
368 accessible surface area (SASA) of the TS, were analyzed using PyMOL (v2.3.0, Schrödinger,
369 LLC). Bottleneck radii of active-site entrances were evaluated using CAVER v3.0[47], and the
370 number of residues mutated to glycine or methionine was counted. Designs were filtered based
371 on the following criteria (Supplementary Table 11): energy difference between bound and
372 unbound structures below 0 kcal mol^{-1}, catalytic residue dihedral angles outside 50° to 130° and
373 –50° to –130°, SASA of TS between 80 and 200 $Å^2$, < 6 Met and < 4 Gly, bottleneck radius of
374 active-site entrance > 0.9 Å, and < 7 non-preorganized residues. Designs passing all criteria were
375 used for structure prediction with AlphaFold2 or ColabFold (v1.3.0), where designs with a
376 backbone RMSD < 3 Å relative to the structure cleaned by Triad, overall pLDDT > 90, and lid
377 pLDDT > 85 were chosen for experimental characterization (Supplementary Table 1).
378
379 ***Sequence optimization of variants 1–5 for enhanced expression.*** Sequence optimization of
380 variants 1–5 was performed using ProteinMPNN with LigandMPNN weights[48]. Amino-acid
381 identities of all residues designed during active-site repacking with Triad were retained. Four
382 sequences were generated using the AlphaFold2 models of each variant as a template, omitting

383 cysteine. The generated sequences were used for structure prediction with ColabFold[45] using all
384 five model weights. Predictions were filtered based on their all-atom RMSD to the input
385 AlphaFold2 model. Theozyme placement was carried out using ColabFold structures, as
386 previously described, to ensure that catalytic contacts were preserved. A single design for each
387 variant was selected for experimental characterization (named KempTIM1–5, Supplementary
388 Table 1).
389
390 ***Ensemble-based design.*** Two ensembles of backbone templates were generated from the
391 KempTIM4 crystal structure (PDB ID: 9QKX) using the ensemble refinement protocol[30]
392 implemented in the PHENIX macromolecular structure determination software. In ensemble
393 refinement, local atomic fluctuations are sampled using molecular dynamics simulations
394 accelerated and restrained by electron density to produce ensemble models fitted to diffraction
395 data. Briefly, input crystal structures were edited to remove low-occupancy conformers and assign
396 an occupancy of 1.0 to the remaining conformer. Following addition of riding hydrogens, parallel
397 ensemble refinement simulations were performed using various combinations of the parameters
398 p_{TLS} (0.6, 0.8, 0.9, 1.0), τ_x (0.5, 1.5, 2.0) and $w_{x\text{-ray}}$ (2.5, 5.0, 10.0), where p_{TLS} describes the
399 percentage of atoms included in a translation-libration-screw (TLS) model use to remove the
400 effects of global disorder, τ_x is the simulation time-step and $w_{x\text{-ray}}$ is the coupled tbath offset, which
401 controls the extent to which the electron density contributes to the simulation force field such that
402 the simulation runs at a target temperature of 300 K. Two ensembles were combined to form a
403 60-member ensemble (Supplementary Table 12).
404 Theozyme placement was performed as described previously on each ensemble member,
405 targeting positions for the base and H-bond donor indicated in Supplementary Table 13.
406 LigandMPNN was used to predict active-site mutations, allowing only hydrophobic and aromatic
407 residues. Amino acids with a mean probability >1% across ten batches were selected as the

408 sequence space for active-site repacking in Triad, performed as described above. Triad models
409 (Supplementary Figure 4) and Boltz-2 predictions (Supplementary Figure 18) were used to filter
410 the designs (Supplementary Table 14 and Supplementary Table 15). Designs were retained only
411 if Boltz-2 models reproduced the intended catalytic contacts and the predicted ligand binding pose
412 deviated by < 1.0 Å root-mean-square deviation from the corresponding Triad model.
413
414 ***Cloning and generation of constructs.*** Genes for all proteins were ordered as codon-optimized
415 fragments flanked by restriction sites for *Nde*I and *Xho*I from Twist Bioscience (South San
416 Francisco, CA, USA). After digestion with *Nde*I and *Xho*I, the gene fragment was ligated into
417 pET21b(+) for variants 1–5 and KempTIM1–5, or pET29b(+) for KempTIM6–9. Catalytic knockout
418 mutations were introduced by a modified QuikChange PCR protocol utilizing KAPA Polymerase
419 followed by *Dpn*I digestion. *E. coli* Top10 or BL21 (DE3) cells were transformed with ligated vector
420 or reaction mixture and plated on Lysogeny Broth (LB) agar plates containing 100 µg mL^{-1}
421 ampicillin or 50 µg mL^{-1} kanamycin as selection markers. Single colonies were used to inoculate
422 5 mL LB supplemented with 100 µg mL^{-1} ampicillin or 50 µg mL^{-1} kanamycin. After overnight
423 growth (37 °C, 250 rpm), cells were harvested by centrifugation and DNA was isolated using
424 NucleoSpin Plasmid EasyPure-Kit (Machery & Nagel) according to the manufacturer's protocol.
425 Vector assembly and introduction of mutations were validated by sequencing (Eurofins
426 Genomics) using standard T7 primers.
427
428 ***Protein expression and purification.*** *E.coli* BL21 (DE3) cells (Novagen) were transformed with
429 plasmid, plated on agar plates containing antibiotic, and incubated overnight at 37 °C. Single
430 colonies were picked to inoculate precultures and incubated at 30–37 °C overnight. On the next
431 day, 1 L LB supplemented with antibiotic was inoculated with 5–10 mL of preculture and incubated
432 at 37 °C until OD600 reached a value between 0.6 and 0.8. Overexpression was induced by

433 adding isopropyl-β-thiogalactoside (IPTG) to a final concentration of 0.1 mM for variants 1–5 and
434 KempTIM1–5 or 1 mM for KempTIM6–9 followed by overnight incubation at 16–20 °C with
435 shaking. Cells were harvested by centrifugation and pellets were either frozen at –20 °C or used
436 directly for purification.

437 Two different purification schemes were employed. In the first scheme, cell pellets were
438 resuspended in 35 mL buffer A (50 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 mM
439 imidazole). Resuspended cells were lysed by sonication (Branson Ultrasonic Sonifier 250, output
440 4, duty cycle 40%, 3 × 3 min) and centrifuged. The supernatant was loaded onto a HisTrapHP
441 column (5 mL, Cytiva Life Science) equilibrated with buffer A and coupled to an ÄKTApure system
442 (Cytiva Life Science). After washing with 10 column volumes (CV) of buffer A, the protein was
443 eluted with a linear gradient over 20 CV to 60 % buffer B (50 mM sodium phosphate pH 7.0, 100
444 mM NaCl, 500 mM imidazole). For crystallisation of KempTIM4-gsg, which contains an additional
445 GSG tripeptide linker between its His-tag and first TIM barrel residue (Supplementary Table 1)
446 and KempTIM1, an additional overnight TEV-cleavage step was performed during a dialysis
447 against buffer C (50 mM of sodium phosphate pH 7.0, 100 mM NaCl). The dialyzed protein was
448 loaded onto a HisTrapHP column (5 mL, Cytiva Life Science) equilibrated with buffer C and
449 coupled to an ÄKTApure system (Cytiva Life Science) and the flowthrough was collected.
450 Fractions containing the protein were pooled, concentrated with a centrifugal concentrator, and
451 loaded onto a HiLoad 26/600 Superdex 75 preparative grade column (Cytiva Life Sciences)
452 preequilibrated in buffer C and eluted with 1 CV buffer C.

453 In the second scheme, cell pellets were resuspended in 8 mL buffer D (5 mM imidazole in
454 100 mM potassium phosphate buffer at pH 8.0) supplemented with 1 mg mL^{-1} lyophilized
455 lysozyme (MP Biomedicals) and 1 U mL^{-1} benzonase nuclease (Merck Millipore). Cells were
456 homogenized using an Avestin EmulsiFlex-B15 cell disruptor and centrifuged. Due to their high
457 thermostability, greater purity for KempTIM4a-h could be achieved through heating the lysate at

458 70 °C for 1 hour prior to centrifugation. Proteins were purified from the lysate by immobilized metal
459 affinity chromatography using Ni-NTA agarose (Qiagen) pre-equilibrated with buffer D in individual
460 Econo-Pac gravity-flow columns (Bio-Rad). Contaminants were washed away using buffer E (10
461 mM imidazole in 100 mM potassium phosphate buffer pH 8.0) and then 20 mM imidazole in the
462 same buffer. Proteins were eluted with 5 mL of buffer F (250 mM imidazole in 100 mM potassium
463 phosphate buffer pH 8.0). Proteins were further subjected to gel filtration in buffer C using an
464 ENrich SEC 70 size-exclusion chromatography column (Bio-Rad). Purity was confirmed by
465 sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and fractions containing
466 the protein of interest were pooled. Protein concentration was determined spectrophotometrically
467 using the absorption at 280 nm and applying Beer-Lambert's law using calculated extinction
468 coefficients obtained from the ExPAsy ProtParam tool (https://web.expasy.org/protparam/).
469
470 ***Far-UV circular dichroism (CD) spectroscopy.*** CD spectra were collected with a Jasco J-710
471 or J-815 spectrometer. Spectra were recorded using a 1-mm quartz cuvette at 20 °C with
472 scanning speed 10–100 nm min^{-1}, bandwidth 1 nm, response time 1 s. Experiments were
473 performed in 20 mM sodium phosphate buffer (pH 7.0), or 50 mM sodium phosphate buffer (pH
474 8.0) supplemented with 100 mM sodium chloride, using a protein concentration of approximately
475 0.2 mg mL^{-1}. For each protein, 3 to 10 spectra were accumulated and averaged. For data
476 normalization, a buffer spectrum was subtracted and the signal was converted to mean residue
477 molar ellipticity using $[\theta_{MRE}] = \theta/(n \times d \times c)$, where n is the number of residues in the protein, θ the
478 collected ellipticity in mdeg, d the path length in mm, and c the protein concentration in M. Thermal
479 denaturation assays were performed by heating the samples from 20 to 95 °C at a rate of 1 °C
480 per minute and ellipticity at 222 nm was measured every 1 °C. Melting temperature (T_m) was
481 determined following the protocol by Greenfield[49].
482

483 ***Size-exclusion chromatography-multi angle light scattering.*** SEC-MALS measurements
484 were performed as described in Beck *et al*.[12] For all measurements, buffer C, an injection volume
485 of 50 µL, and a protein concentration of 2 mg mL^{-1} were used.
486
487 ***Steady-state kinetics.*** Assays were performed in sodium phosphate (pH 7.0), Tris (pH 7.0–8.5),
488 or CHES (pH 9.0–10.0) buffers, all at 50 mM with 100 mM NaCl and 10% MeOH in a Spark
489 (Tecan) or Synergy H1 (Biotek) plate reader. Enzyme concentrations varied from 0.2 to 80 µM.
490 Reactions (200 µL final volume) were initiated by addition of varying concentrations (0.015–2 mM)
491 of 5-nitrobenzisoxazole (abcr or AAblocks) or 6-nitrobenzisoxazole (AAblocks) dissolved in
492 methanol (final methanol concentration 10%). Product formation was monitored at 380 nm
493 (ε = 15,800 M^{-1} cm^{-1})[6] or 400 nm (ε = 2,870 M^{-1} cm^{-1})[50] for 5-nitrobenzisoxazole and 6-
494 nitrobenzisoxazole, respectively, in individual wells of a 96-well plate (Nunc or Greiner Bio-One).
495 Path lengths for each well were calculated ratiometrically using the difference in absorbance at
496 900 and 975 nm. Control measurements without protein were conducted for each substrate
497 concentration and subtracted from the enzyme measurements. Linear phases of the kinetic traces
498 were used to measure initial reaction rates. If the enzyme showed saturation within the substrate
499 concentration used, the data were fitted to the Michealis-Menten model using
500 ($v_0 = (v_{max} \times [S]) / (K_M + [S])$ where v_0 is the initial velocity rate, v_{max} is the maximal velocity rate,
501 [S] is the substrate concentration, K_M is the Michaelis constant. The catalytic constant k_{cat} was
502 calculated with $k_{cat} = v_{max} / [E]$ where [E] is the enzyme concentration. If saturation was not
503 achieved within the substrate's solubility limit, a linear equation ($v_0 = k_{cat}/K_M \times [S]$) was used.
504
505 ***Crystallization and structure determination.*** KempTIM4-gsg and KempTIM1 (Supplementary
506 Table 1) were purified using the first purification scheme, with the final size-exclusion
507 chromatography step performed in 20 mM HEPES buffer pH 8.0 supplemented with 20 mM NaCl.

508 Crystallization screens were set up using the sitting-drop vapor diffusion method with a Phoenix
509 pipetting robot (Art Robbins Instruments) and commercially available sparse-matrix screens
510 (NeXtal) in 96-well sitting-drop plates (3-drop Intelli-Plates, Art Robbins Instruments). Protein,
511 reservoir solutions were mixed in ratios of 1:1, 2:1, and 1:2 and crystals were grown at 293 K.
512 The protein concentration of KempTIM4-gsg was 12 mg mL^{-1}, supplemented with 5 mM 6NBT
513 (5% DMSO), and diffracting crystals were obtained after 40 days in 4 M sodium formate.
514 Cryoprotection was achieved by adding glycerol to a final concentration of 25%.

515 The protein concentration of KempTIM1 was 8.75 mg mL^{-1}. For the inhibitor-bound
516 complex structure 6NBT was added as a solid in excess and the solution was incubated for 1 h
517 and subsequently centrifuged. Apo KempTIM1 crystals were obtained after 48 h in droplets
518 containing 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 5.5, 25% w/v PEG 3350 as reservoir
519 solution. The 6NBT complex crystals were observed after 24 h in droplets with 0.1 M sodium
520 acetate pH 4.5, 0.1 M Bis-Tris pH 5.5, 25% w/v PEG 3350 as reservoir. The KempTIM1 crystals
521 were cryo-protected by adding 20% ethylene glycol to the corresponding reservoir solution. All
522 crystals were mounted using cryo-loops on SPINE standard bases and flash-cooled in liquid
523 nitrogen.

524 Diffraction data were collected at 100 K at beamline BL 14.1 with a Pilatus3 S 6M detector
525 at the BESSY II synchrotron (Helmholtz-Zentrum Berlin). Data processing was performed using
526 X-ray Detector Software APP3 (XDSAPP3)[51] with XDS[52], and data quality was assessed using
527 phenix.xtriage[53]. Phases were solved by molecular replacement using AlphaFold2 predictions as
528 search models with Phaser[54]. The resulting models were manually rebuilt using Coot[55] and refined
529 with phenix.refine[53] in an iterative process.

530 The structure of KempTIM4-gsg was refined to a resolution of 2.3 Å. Despite crystallization
531 with 6NBT, no electron density corresponding to the transition-state analogue was observed in
532 the active site; instead, the density was best interpreted as a bound glycerol molecule from the

533 cryoprotectant. The structures of KempTIM1 apo and co-crystallized with 6NBT were refined to
534 resolutions of 1.25 Å and 1.20 Å, respectively. The high resolution of the co-crystal structure and
535 the well-defined electron density of 6NBT in the active site enabled conclusive placement of the
536 transition-state analogue. Additionally, there is also a diffuse electron density, which cannot be
537 interpreted unambiguously, indicating weak binding of solvent molecules. However, a putative
538 ethylene glycerol molecule was placed into the density to account for the residual density, as it
539 was used as cryoprotectant. Refinement statistics and crystallographic data are shown in
540 Supplementary Table 6. Coordinates and structure factors were validated and deposited in the
541 PDB under accession codes 9TZD, 9U01 and 9QKX.
542
543 ***Size-exclusion chromatography small-angle X-ray scattering (SEC-SAXS).*** SEC-SAXS
544 measurements were conducted at the BioSAXS beamline BM29 of the ESRF in Grenoble,
545 France. A Superdex 75 Increase 10/300 GL column (Cytiva Life Sciences) was used with a flow
546 rate of 0.8 mL min^{-1} in buffer C. KempTIM4-gsg was measured at a concentration of 5 mg mL^{-1},
547 with an injection volume of 100 µL. Data processing and analysis were performed using ATSAS
548 3.2.1[56,57]. For analysis, AlphaFold2 models of KempTIM4 and NT6-CP2, as well as an
549 AlphaFlow[58] model of KempTIM4 with a displaced lid obtained using the base molecular dynamics
550 weights were used. For each of these models a theoretical scattering curve were generated and
551 fitted to the experimental data using CRYSOL[59].
552
553 ***Molecular dynamics (MD) simulations.*** The Amber 2020 software (http://ambermd.org/) with
554 the ff19SB protein force field[60], gaff2 ligand force field[61], and OPC water model[62] was used for all
555 simulations. A cutoff of 10 Å was applied to electrostatics modeled using the particle mesh Ewald
556 method[63]. All MD trajectories were run with a time step of 2 fs. Models of each design generated
557 using Triad as described above were prepared for MD. MD trajectories were run for each enzyme

558 in the presence and absence of bound transition-state analogue 6-nitrobenzotriazole (6NBT).
559 Parameters for 6NBT were generated using the Antechamber package[64]. Hydrogen atoms were
560 added using Reduce[65], and the prepared structures were solvated with OPC water in a truncated
561 octahedral box with periodic boundary conditions where the distance between the protein surface
562 and the box edges was set to 10 Å. The addions2 algorithm in Amber was used to place
563 counterions to neutralize the system. The system was minimized (solvent, protein, then whole
564 system) using the steepest descent method with positional restraints of 500 kcal mol^{-1} $Å^{-2}$. The
565 system was then heated to a temperature of 300 K over 240 ps while restraints were gradually
566 removed, followed by equilibration under an NPT ensemble at 300 K and 1 bar for 10 ns and
567 subsequent equilibration under an NVT ensemble at 300 K for another 10 ns. Constant
568 temperature and pressure were achieved using the Langevin thermostat[66] and Berendsen
569 barostat[67], respectively. Following equilibration, ten replicas of 100-ns production simulations
570 were run. For analyzing catalytic contacts, 1000 snapshots separated by 0.1 ns each were
571 extracted from the production trajectory. Catalytic interactions and RMSF values were extracted
572 using CPPTRAJ and PYTRAJ, respectively[68].

573 **Table 1. Kinetic parameters of selected KempTIMs**

Enzyme	Base	H-bond donor	pH	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
KempTIM1	D24	N131	7	4.2 ± 0.2	3.1 ± 0.2	1,400 ± 100
			10	13.7 ± 0.5	0.66 ± 0.06	21,000 ± 2,000
	D24A	N131	7	N.D.	N.D.	N.D.
	D24	N131A	7	0.65 ± 0.02	2.6 ± 0.1	250 ± 10
KempTIM4	D153	N46	7	N.D.	N.D.	7.01 ± 0.08
			10	N.D.	N.D.	18.6 ± 0.4
	D153A	N46	7	N.D.	N.D.	N.D.
	D153	N46A	7	N.D.	N.D.	6.5 ± 0.1
KempTIM4b	D153	Q42	7	1.21 ± 0.04	0.39 ± 0.04	3,100 ± 300
			10	26.3 ± 0.9	0.83 ± 0.06	32,000 ± 3,000
	D153A	Q42	7	N.D.	N.D.	N.D.
	D153	Q42A	7	0.46 ± 0.03	0.7 ± 0.1	640 ± 90

574 k_{cat} and K_M using 5-nitrobenzisoxazole as substrate were calculated by fitting the data to the Michaelis-Menten model $v_0 =$
575 $k_{cat}[E_0][S]/(K_M+[S])$. Errors of nonlinear regression fitting, which represent the absolute measure of the typical distance that each data
576 point falls from the regression line, are provided. N.D. indicates that no activity was detected, or that K_M and k_{cat} could not be
577 determined accurately because saturation was not possible at the maximum concentration of 5-nitrobenzisoxazole tested (2 mM),
578 which is the substrate's solubility limit. In those cases, catalytic efficiencies (k_{cat}/K_M) were calculated from the slope of the linear
579 portion ($[S] \ll K_M$) of the Michaelis-Menten model ($v = (k_{cat}/K_M)[E_0][S]$). n = 2 independent protein batches for KempTIM1 and
580 KempTIM4b. n = 4 independent protein batches for KempTIM4. At least two technical replicates were performed for each
581 independent batch. Kinetic assays were carried out in 50 mM sodium phosphate containing 100 mM NaCl and 10% MeOH (pH 7) or
582 50 mM CHES containing 100 mM NaCl and 10% MeOH (pH 10).
583
584

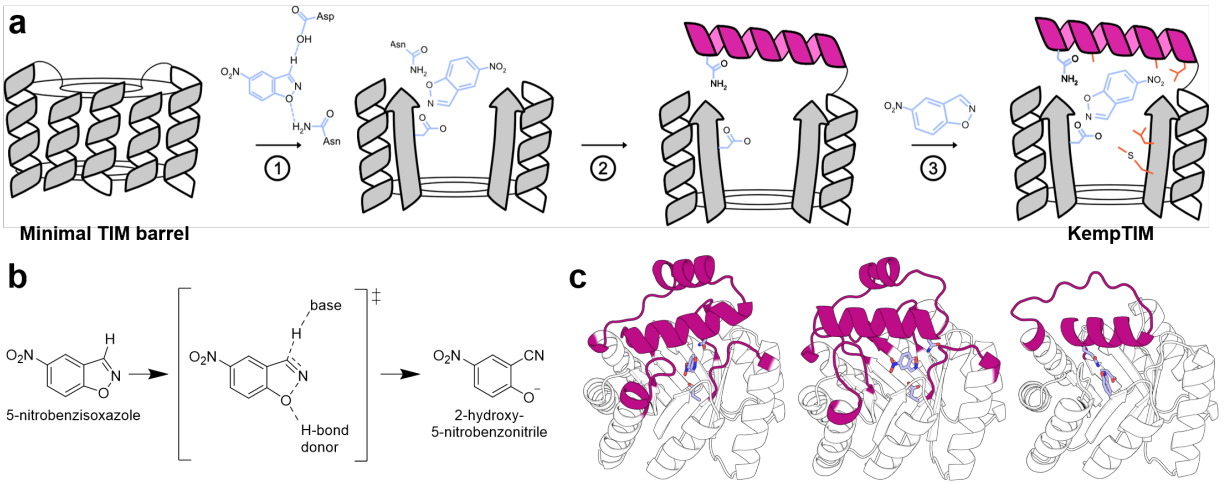

Figure 1. *De novo* enzyme design using CANVAS. (a) CANVAS follows three general steps: (1) placement of the theozyme (blue) onto the minimal TIM-barrel scaffold using a single catalytic residue as an anchor; (2) construction of a custom lid (magenta) to form the active-site pocket and anchor the second catalytic residue; and (3) active-site repacking to further stabilize the transition state. (b) Kemp elimination reaction. (c) Structure of three distinct lids (magenta) tailored to the Kemp elimination, designed using CANVAS. Transition state and designed catalytic residues are shown as sticks.

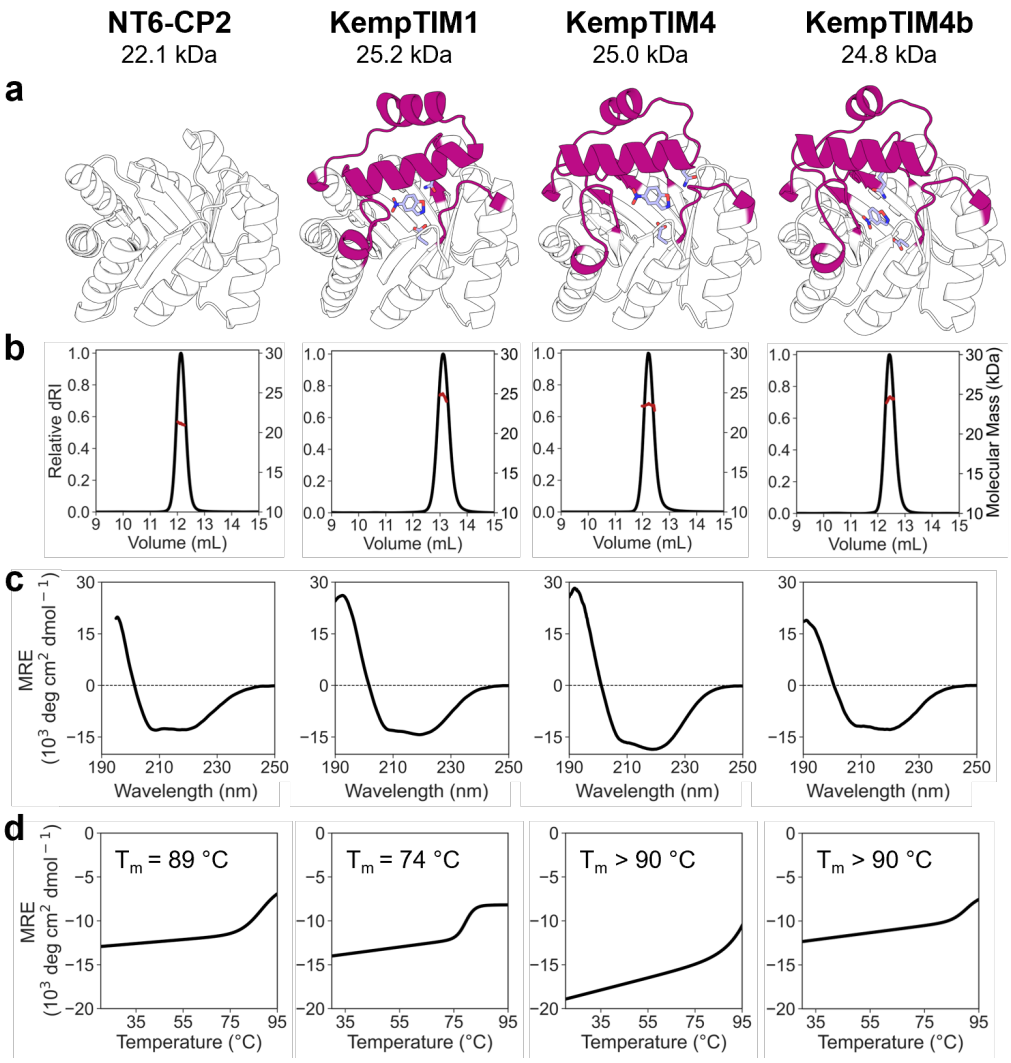

Figure 2. Structural characterization. (a) Computational models of minimal TIM-barrel NT6-CP2 and KempTIMs with the designed lid shown in magenta and the theozyme as sticks. (b) SEC-MALS indicates that the proteins are monomeric, with molecular weights close to their theoretical values. dRI: differential refractive index. (c) Circular dichroism spectra reveal a mixed αβ signal characteristic of TIM barrels. MRE: mean residue ellipticity. (d) Melting curves demonstrate that KempTIM1 is substantially destabilized compared to NT6-CP2. T_m: melting temperature.

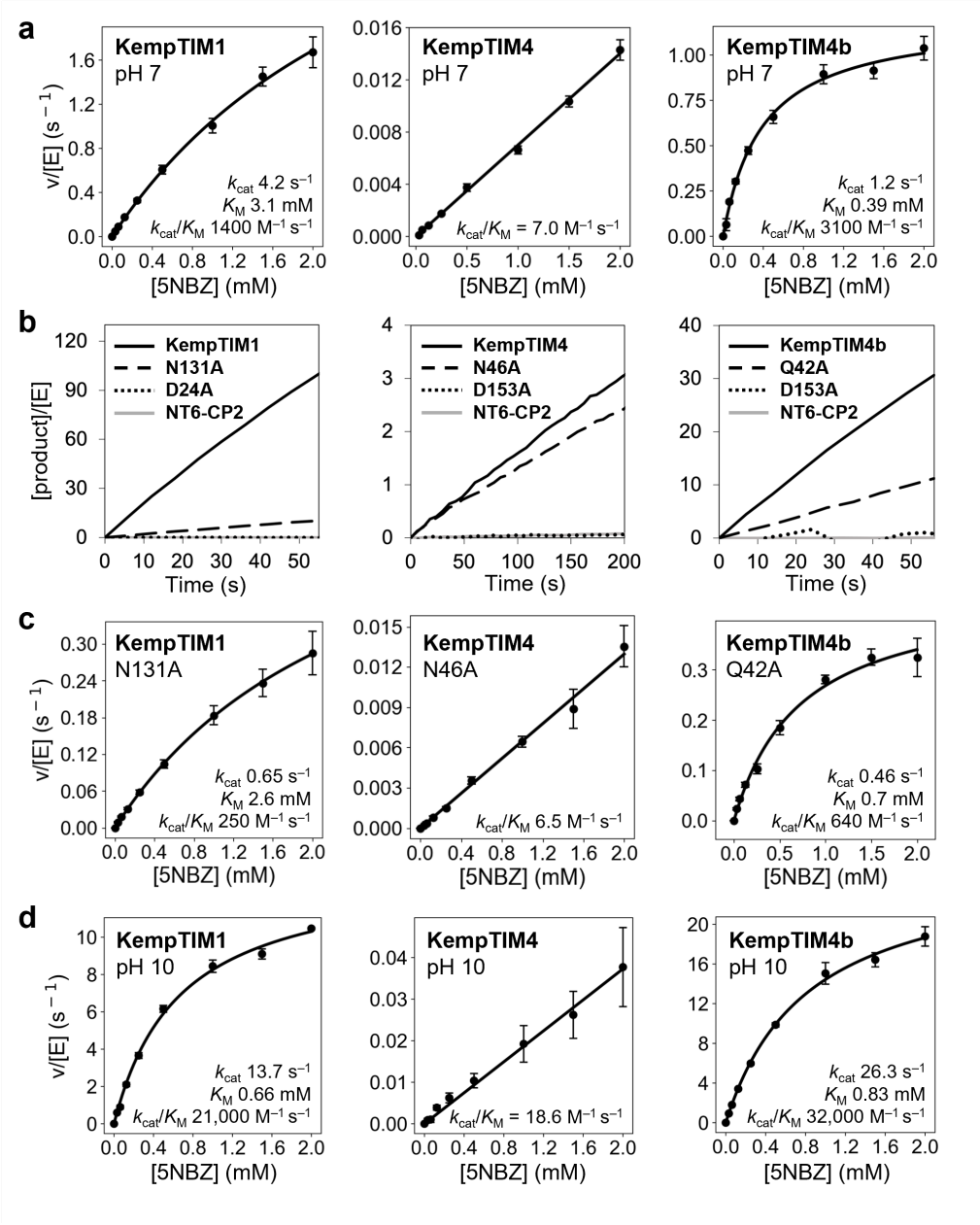

Figure 3. Kinetic characterization. Assays were carried out in 50 mM sodium phosphate containing 100 mM NaCl and 10% MeOH (pH 7) or 50 mM CHES containing 100 mM NaCl and 10% MeOH (pH 10). (a) Michaelis-Menten plots (pH 7) showing normalized initial rates as a function of 5-nitrobenzisoxazole (5NBZ) concentration. Data represent the mean ± SEM of ≥ 5 replicate measurements from ≥ 2 independent protein batches. (b) Reaction progress curves (pH 7) at 2 mM 5NBZ illustrating the contributions of the designed catalytic residues to overall catalysis. (c) Michaelis-Menten plots (pH 7) for variants lacking H-bond donor residues. Data represent the mean ± s.d. of ≥ 3 replicate measurements from one independent protein batch. (d) Michaelis-Menten plots (pH 10) showing normalized initial rates as a function of 5NBZ concentration. Data represent the mean ± SEM of ≥ 4 replicate measurements from ≥ 2 independent protein batches.

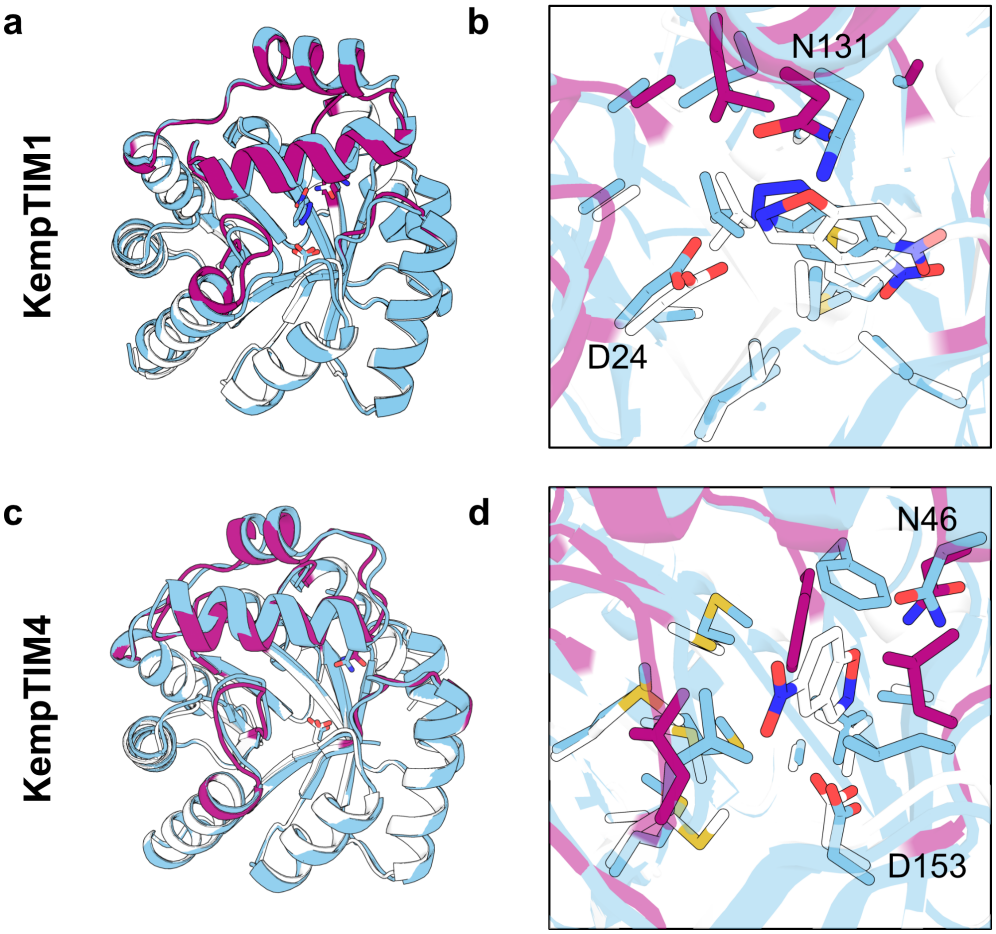

Figure 4. Crystal structures. Experimental structures are shown in blue and design models in white (minimal TIM barrel) and magenta (lid). (a) The crystal structure of KempTIM1 closely matches the design model, with backbone Cα root-mean-square deviations (RMSDs) of 0.68 Å for the full structure and 0.88 Å for the lid. (b) Active-site residues within the minimal TIM barrel of KempTIM1 more closely match their designed conformations than those in the lid, consistent with an ~1 Å displacement of the lid relative to the TIM-barrel core. The co-crystallized transition-state analogue 6NBT adopts a binding pose similar to that in the design model, enabling hydrogen bonding to both the catalytic base D24 and the H-bond donor N131, despite N131 adopting a different rotamer in the crystal structure. (c) The crystal structure of KempTIM4 also shows close agreement with the design model (backbone Cα RMSD = 0.86 Å for the full structure and 1.34 Å for the lid). (d) Active-site residues of KempTIM4 show substantial divergence from the design model, potentially reflecting the absence of a bound transition-state analogue. The catalytic base D153 adopts a similar rotamer in both structures, whereas the catalytic H-bond donor N46 adopts a rotamer incompatible with transition-state stabilization, likely caused by the ~2 Å displacement of the lid relative to the TIM-barrel core.

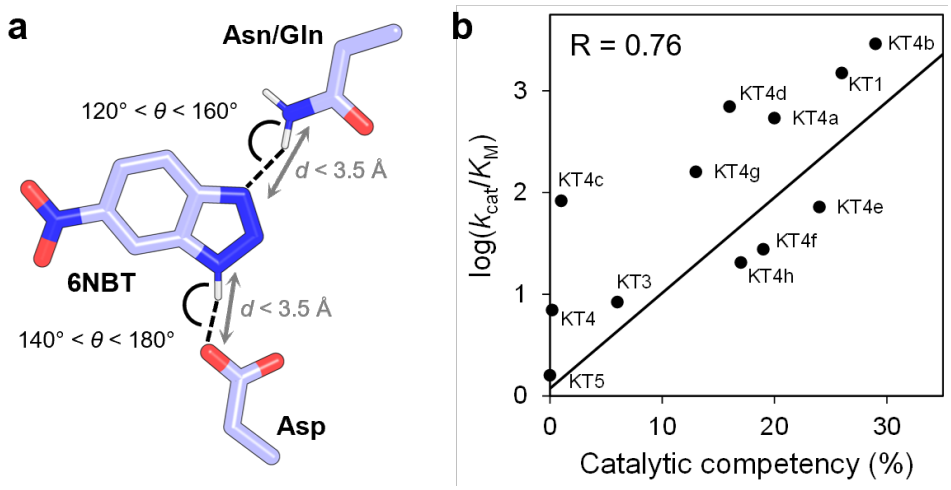

Figure 5. MD-derived catalytic competency correlates with experimental activity. (a) Geometric definitions of the catalytic contacts between the transition-state analogue (6NBT) and the designed general base (Asp) and H-bond donor (Asn/Gln). Catalytic angle (θ) and heavy-atom distance (d) thresholds defining competent contacts are indicated. (b) Correlation between catalytic-competency score, defined as the fraction of MD snapshots in which both catalytic contacts simultaneously satisfy the geometric criteria, and experimental catalytic efficiencies (k_{cat}/K_M). KempTIM (KT) variants with higher catalytic competency display higher activities (R = 0.76), with KempTIM1 and KempTIM4b ranking highest.

Acknowledgments

J.B. and B.H. acknowledge support from the Elite Network of Bavaria and its study program "Biological Physics". B.H. acknowledges support from European Union's Horizon 2020 research and innovation program under grant agreement No 951375 (ArtMotor). R.A.C. acknowledges grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2021-03484 and RGPAS-2021-00017) and the Canada Foundation for Innovation (26503). B.J.S. was supported by an NSERC CREATE scholarship. This research was enabled in part by support provided by Compute Ontario (www.computeontario.ca) and the Digital Research Alliance of Canada (alliancecan.ca). We acknowledge financial support and allocation of synchrotron beamtime by HZB at BESSY and at ESRF in Grenoble and thank the beamline staff for support. We thank Sabrina Wischt for technical support, Janosch Hennig for help with SAXS data analysis and the University of Bayreuth Centre of International Excellence "Alexander von Humboldt" for facilitating this collaboration through a short-term grant to R.A.C.

Competing interests

The authors declare no competing financial interest.

ORCID identifiers

Julian Beck: 0009-0007-3555-9890

Roberto A. Chica: 0000-0003-3789-9841

Emily Freund: 0009-0006-2876-3319

Ahana G. Harsha: 0009-0001-2150-0042

Jan Hartmann: 0009-0008-7160-0034

Birte Höcker: 0000-0002-8250-9462

Mark Kriegel: 0000-0002-0489-0923

Benjamin J. Smith: 0009-0003-5953-4998

Niayesh Zarifi: 0000-0003-4748-7082

References

1. Nagano, N., Orengo, C. A. & Thornton, J. M. One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on their Sequences, Structures and Functions. *Journal of Molecular Biology* **321**, 741–765 (2002).
2. Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. Stability, catalytic versatility and evolution of the (βα)8-barrel fold. *Current Opinion in Biotechnology* **12**, 376–381 (2001).
3. Wierenga, R. K., Kapetaniou, E. G. & Venkatesan, R. Triosephosphate isomerase: a highly evolved biocatalyst. *Cellular and Molecular Life Sciences* **67**, 3961–3982 (2010).
4. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
5. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–91 (2008).
6. Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3790–5 (2012).
7. Romero-Romero, S., Kordes, S., Michel, F. & Höcker, B. Evolution, folding, and design of TIM barrels and related proteins. *Current Opinion in Structural Biology* **68**, 94–104 (2021).
8. Huang, P.-S. *et al.* De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology* **12**, 29–34 (2016).
9. Kordes, S., Romero-Romero, S., Lutz, L. & Höcker, B. A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. *Protein Science* **31**, 513–527 (2022).
10. Romero-Romero, S. *et al.* The Stability Landscape of de novo TIM Barrels Explored by a Modular Design Approach. *Journal of Molecular Biology* **433**, 167153 (2021).

688 11. Chu, A. E., Fernandez, D., Liu, J., Eguchi, R. R. & Huang, P.-S. De Novo Design of a
689 Highly Stable Ovoid TIM Barrel: Unlocking Pocket Shape towards Functional Design.
690 *BioDesign Research* **2022**, 9842315 (2022).
691 12. Beck, J., Shanmugaratnam, S. & Höcker, B. Diversifying de novo TIM barrels by
692 hallucination. *Protein Science* **33**, e5001 (2024).
693 13. Kordes, S., Beck, J., Shanmugaratnam, S., Flecks, M. & Höcker, B. Physics-based
694 approach to extend a de novo TIM barrel with rationally designed helix-loop-helix motifs.
695 *Protein Engineering, Design and Selection* **36**, gzad012 (2023).
696 14. Wiese, J. G., Shanmugaratnam, S. & Höcker, B. Extension of a de novo TIM barrel with
697 a rationally designed secondary structure element. *Protein Science* **30**, 982–989 (2021).
698 15. Caldwell, S. J. *et al.* Tight and specific lanthanide binding in a de novo TIM barrel with a
699 large internal cavity designed by symmetric domain fusion. *Proceedings of the National*
700 *Academy of Sciences* **117**, 30362–30369 (2020).
701 16. Klein, A. S. *et al.* A De Novo Metalloenzyme for Cerium Photoredox Catalysis. *J. Am.*
702 *Chem. Soc.* **146**, 25976–25985 (2024).
703 17. Tantillo, D. J., Jiangang, C. & Houk, K. N. Theozymes and compuzymes: theoretical
704 models for biological catalysis. *Current Opinion in Chemical Biology* **2**, 743–750 (1998).
705 18. Lee, F. S., Anderson, A. G. & Olafson, B. D. Benchmarking TriadAb using targets from
706 the second antibody modeling assessment. *Protein Engineering, Design and Selection* **36**,
707 gzad013 (2023).
708 19. Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. Combinatorial methods for small-
709 molecule placement in computational enzyme design. *Proceedings of the National Academy*
710 *of Sciences of the United States of America* **103**, 16710–5 (2006).
711 20. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion.
712 *Nature* **620**, 1089–1100 (2023).

713 21. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using
714 ProteinMPNN. *Science* **378**, 49–56 (2022).
715 22. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
716 583–589 (2021).
717 23. Rakotoharisoa, R. V. *et al.* Design of Efficient Artificial Enzymes Using
718 Crystallographically Enhanced Conformational Sampling. *Journal of the American Chemical*
719 *Society* **146**, 10001–10013 (2024).
720 24. Anand, N. *et al.* Protein sequence design with a learned potential. *Nature*
721 *Communications* **13**, 746 (2022).
722 25. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**,
723 387–394 (2022).
724 26. Sumida, K. H. *et al.* Improving Protein Expression, Stability, and Function with
725 ProteinMPNN. *Journal of the American Chemical Society* **146**, 2054–2061 (2024).
726 27. Listov, D. *et al.* Complete computational design of high-efficiency Kemp elimination
727 enzymes. *Nature* **643**, 1421–1427 (2025).
728 28. Chen, Y. *et al.* Emergence of specific binding and catalysis from a designed generalist
729 binding protein. *bioRxiv* 2025.01.30.635804 (2025) doi:10.1101/2025.01.30.635804.
730 29. Broom, A. *et al.* Ensemble-based enzyme design can recapitulate the effects of
731 laboratory directed evolution in silico. *Nature Communications* **11**, 4808 (2020).
732 30. Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein
733 crystal structures by ensemble refinement. *eLife* **1**, e00311 (2012).
734 31. Passaro, S. *et al.* Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction.
735 *bioRxiv* 2025.06.14.659707 (2025) doi:10.1101/2025.06.14.659707.
736 32. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp
737 eliminase. *Nature* **503**, 418–21 (2013).

738 33. Sterner, R. & Höcker, B. Catalytic Versatility, Stability, and Evolution of the (βα)8-Barrel
739 Enzyme Fold. *Chemical Reviews* **105**, 4038–4055 (2005).
740 34. Zhang, S., Zhang, J., Luo, W., Wang, P. & Zhu, Y. A preorganization oriented
741 computational method for de novo design of Kemp elimination enzymes. *Enzyme and*
742 *Microbial Technology* **160**, 110093 (2022).
743 35. Bhattacharya, S. *et al.* NMR-guided directed evolution. *Nature* **610**, 389–393 (2022).
744 36. Braun, M. *et al.* Computational enzyme design by catalytic motif scaffolding. *Nature*
745 https://doi.org/10.1038/s41586-025-09747-9 (2025) doi:10.1038/s41586-025-09747-9.
746 37. Lauko, A. *et al.* Computational design of serine hydrolases. *Science* **388**, eadu2454
747 (2025).
748 38. Koga, N. *et al.* Role of backbone strain in de novo design of complex α/β protein
749 structures. *Nature Communications* **12**, 3921 (2021).
750 39. Koga, N. & Tatsumi-Koga, R. Inventing Novel Protein Folds. *Journal of Molecular*
751 *Biology* **436**, 168791 (2024).
752 40. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation
753 and design of macromolecules. *Methods Enzymol* **487**, 545–74 (2011).
754 41. Dunbrack, R. L. Rotamer Libraries in the 21st Century. *Current Opinion in Structural*
755 *Biology* **12**, 431–440 (2002).
756 42. Chica, R. A., Moore, M. M., Allen, B. D. & Mayo, S. L. Generation of longer emission
757 wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of*
758 *the National Academy of Sciences of the United States of America* **107**, 20257–62 (2010).
759 43. Mayo, S. L., Olafson, B. D. & Goddard, W. A. Dreiding - a Generic Force-Field for
760 Molecular Simulations. *Journal of Physical Chemistry* **94**, 8897–8909 (1990).

761 44. Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in protein design.
762 *Proceedings of the National Academy of Sciences of the United States of America* **94**,
763 10172–7 (1997).
764 45. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**,
765 679–682 (2022).
766 46. Allen, B. D., Nisthal, A. & Mayo, S. L. Experimental library screening demonstrates the
767 successful application of computational protein design to large structural ensembles.
768 *Proceedings of the National Academy of Sciences of the United States of America* **107**,
769 19838–43 (2010).
770 47. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in
771 Dynamic Protein Structures. *PLOS Computational Biology* **8**, e1002708 (2012).
772 48. Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using
773 LigandMPNN. *Nature Methods* **22**, 717–723 (2025).
774 49. Greenfield, N. J. Determination of the folding of proteins as a function of denaturants,
775 osmolytes or ligands using circular dichroism. *Nature Protocols* **1**, 2733–2741 (2006).
776 50. Bunzel, H. A. *et al.* Emergence of a Negative Activation Heat Capacity during Evolution
777 of a Designed Enzyme. *Journal of the American Chemical Society* **141**, 11745–11748 (2019).
778 51. Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. XDSAPP2.0. *Journal*
779 *of applied crystallography* **49**, 1085–1092 (2016).
780 52. Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125–132 (2010).
781 53. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
782 electrons: recent developments in Phenix. *Acta Crystallographica Section D* **75**, 861–877
783 (2019).
784 54. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography*
785 **40**, 658–674 (2007).

786 55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
787 *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501 (2010).
788 56. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: improvements to a free open-
789 source program for small-angle X-ray scattering data reduction and analysis. *Journal of*
790 *applied crystallography* **50**, 1545–1553 (2017).
791 57. Manalastas-Cantos, K. *et al.* ATSAS 3.0: expanded functionality and new tools for small-
792 angle scattering data analysis. *Journal of applied crystallography* **54**, 343–355 (2021).
793 58. Jing, B. AlphaFold Meets Flow Matching for Generating Protein Ensembles. *arXiv* **2402**,
794 04845 (2024).
795 59. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle
796 scattering from macromolecular solutions. *Journal of applied crystallography* **50**, 1212–1225
797 (2017).
798 60. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB
799 protein force field. *Proteins: Structure, Function, and Bioinformatics* **78**, 1950–1958 (2010).
800 61. He, X., Man, V. H., Yang, W., Lee, T.-S. & Wang, J. A fast and high-quality charge
801 model for the next generation general AMBER force field. *The Journal of Chemical Physics*
802 **153**, 114502 (2020).
803 62. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different
804 Approach. *The Journal of Physical Chemistry Letters* **5**, 3863–3871 (2014).
805 63. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald
806 sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).
807 64. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type
808 perception in molecular mechanical calculations. *Journal of Molecular Graphics and*
809 *Modelling* **25**, 247–260 (2006).

810 65. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and
811 glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation.
812 *Journal of Molecular Biology* **285**, 1735–47 (1999).
813 66. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of
814 molecular dynamics. *The Journal of Chemical Physics* **114**, 2090–2098 (2001).
815 67. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R.
816 Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**,
817 3684–3690 (1984).
818 68. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and
819 Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and*
820 *Computation* **9**, 3084–3095 (2013).
821