Customizing the Structure of a Minimal TIM Barrel to Craft a De Novo Enzyme

Julian Beck,[1,†] Benjamin J. Smith,[2,3,†] Niayesh Zarifi,[2,3] Emily Freund,[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 architectures lack 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. Experimental testing identified one active variant with catalytic efficiency comparable to previously reported Kemp eliminases, and mutational analyses validated the essential role of the designed catalytic residues. Sequence optimization of this variant improved solubility and stability, enabling X-ray structure determination, which confirmed the designed lid structure. This study reports the first enzymatically active *de novo* TIM barrel and establishes a platform for designing enzymes 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]. While idealized *de novo* TIM barrels have been successfully designed[4-7], none have been converted into functional enzymes 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[8-10]. However, these pockets are too solvent exposed and undefined to support the microenvironments required for catalysis. Therefore, novel strategies are needed to craft custom active sites and endow *de novo* TIM barrels with 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[11]. Using the protein design software Triad[12], 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[13]. A second catalytic residue is then built from the transition state, ensuring proper geometry for catalytic interaction. The alpha carbon of this second residue is thus positioned in the empty space above the TIM barrel catalytic face. This allows it to serve as an anchor for constructing a custom lid to form the active-site pocket, which is created using the generative AI tool RFdiffusion[14] followed by sequence design with ProteinMPNN[15] and evaluation with AlphaFold2[16]. The active site is further optimized with Triad to build a pocket complementary to the transition state[17]. Designs are then filtered using key enzyme design criteria[18], including

58 solvent-accessible surface area of the transition state, active-site preorganization, energy, and
59 catalytic contact geometry.
60
61 **Applying CANVAS to minimal TIM-barrel proteins**
62 We applied CANVAS to the Kemp elimination (Figure 1b), a model organic transformation
63 commonly used as a benchmark for *de novo* enzyme design[17-19]. The reaction's theozyme
64 features an Asp as catalytic base and an Asn or Gln as H-bond donor to stabilize negative charge
65 buildup on the oxygen at the transition state. We used the crystal structures of two *de novo* TIM
66 barrels (PDB IDs: 7MCD[20] and 7OSV[5]) as templates. Since 7OSV has its N- and C-termini on the
67 catalytic face of the TIM barrel, we applied an *in silico* circular permutation based on the inpainting
68 method[21] to generate two modified structures termed NT6-CP1 and NT6-CP2 that contain termini
69 on the stability face (Supplementary Figure 2). Starting from these structures, we used CANVAS
70 to generate a distinct lid for each template tailored to the Kemp elimination (Figure 1c). The lid
71 diffused onto 7MCD consists of a single 26-residue fragment introduced in βα loop 7
72 (Supplementary Figure 3). By contrast, the designed lids in NT6-CP1 and NT6-CP2 comprise a
73 major fragment of 33 or 39 residues introduced into βα loops 2 or 6, respectively, to scaffold the
74 catalytic H-bond donor, as well as three or four elongated loops to enhance interactions with the
75 primary fragment. The resulting designs, termed KempTIMs, include variants 1–3 obtained from
76 NT6-CP2, variants 4–5 from NT6-CP1 and variants 6–9 from 7MCD, each with a unique active-
77 site configuration (Supplementary Figure 4) and sequence (Supplementary Table 1), which
78 yielded high AlphaFold2 pLDDT scores (Supplementary Figure 5).
79
80 **Kinetic assays reveal an active Kemp eliminase**
81 Of the nine designs, seven expressed in soluble form in *E. coli* but only five could be
82 purified to homogeneity (Supplementary Table 2). Notably, the NT6-CP1 template could not be

83 expressed, in contrast to NT6-CP2 and 7MCD. Enzymatic assays revealed a single active design,
84 KempTIM4, which followed Michaelis-Menten kinetics with substrate saturation (Figure 2a),
85 yielding k_{cat}, K_M, and k_{cat}/K_M values of 0.0066 ± 0.0006 s^{-1}, 0.6 ± 0.1 mM, and 11 ± 3 M^{-1} s^{-1},
86 respectively. This catalytic efficiency is comparable to those of other *de novo* Kemp
87 eliminases[17,19,22], such as KE07 (k_{cat} = 0.018 s^{-1}, K_M = 1.4 mM, and k_{cat}/K_M = 12.2 M^{-1} s^{-1})[19].
88 KempTIM4 expressed predominantly as a monomeric protein (Figure 3a,b, Supplementary Table
89 3) and displayed a mixed αβ secondary structure characteristic of TIM barrels (Figure 3c).
90 Compared to the minimal TIM barrel NT6-CP2, KempTIM4 exhibited increased random coil
91 content, indicated by a stronger circular dichroism (CD) signal at 208 nm. This structural feature
92 aligns with the design of its lid, which includes long loops (Figure 3a, Supplementary Figure 3).
93 KempTIM4 showed a strongly reduced melting temperature relative to NT6-CP2 (Figure 3d),
94 reflecting the destabilizing effects of incorporating a lid and active site onto the minimal TIM barrel
95 scaffold. Mutation of the designed catalytic base and H-bond donor to Ala abolished or
96 substantially reduced activity (Figure 2b) despite size-exclusion chromatograms and CD spectra
97 comparable to KempTIM4 (Supplementary Figure 6, Supplementary Table 3). These results
98 confirm that the loss of activity in these knockout variants is due to mutation, not structural
99 disruption. Moreover, the absence of enzymatic activity in NT6-CP2 demonstrates that catalysis
100 in KempTIM4 is driven by the designed lid and active site.
101
102 **Redesign yields a more stable enzyme**
103 Although active, KempTIM4 was prone to aggregation at high concentrations. This led us
104 to redesign its sequence for enhanced expression and solubility using ProteinMPNN, which has
105 been shown to produce stable and soluble proteins[23]. In this procedure, identities of active-site
106 residues designed with Triad were preserved to maintain catalytic function, while the rest of the
107 protein was redesigned. The redesigned protein, KempTIM4-2, demonstrated approximately 10-

108 fold higher expression yields compared to KempTIM4 (Supplementary Table 2) and existed
109 exclusively as a monomer with no tendency to form dimers (Figure 3b, Supplementary Table 3).
110 CD spectroscopy showed increased α-helicity compared to KempTIM4 and higher stability,
111 comparable to NT6-CP2 (Figure 3c,d). Despite these improvements, KempTIM4-2 exhibited lower
112 activity than KempTIM4 (k_{cat}/K_M = 7.0 ± 0.1 M^{-1} s^{-1}) and could not be saturated within the
113 substrate's solubility limit (Figure 2a), even though active-site residues remained unchanged. As
114 with KempTIM4, mutation of the catalytic base in KempTIM4-2 abolished activity (Figure 2b)
115 without disrupting structure (Supplementary Figure 6, Supplementary Table 3). However,
116 mutation of the H-bond donor caused only a small reduction in catalytic activity compared to
117 KempTIM4, suggesting that this designed Asn residue adopts a suboptimal orientation hindering
118 its catalytic function. Nevertheless, inhibition assays using the 6-nitrobenzotriazole (6NBT)
119 transition-state analogue demonstrated competitive binding in the active site (Figure 2c),
120 confirming the integrity of the designed active site pocket.
121
122 **The crystal structure confirms the designed lid and provides active site details**
123 Crystals of KempTIM4-2 were obtained in the presence of 6NBT, and the structure was
124 solved at 2.3 Å resolution (Supplementary Table 4). The structure confirmed the correct folding
125 of the designed lid, with the primary helix-turn-helix fragment and four elongated loops closely
126 matching the design model (Figure 4a). However, the lid was slightly displaced relative to the TIM
127 barrel core, resulting in a 1.8-Å shift of the catalytic H-bond donor N46 Cα atom (Figure 4b), which
128 likely reduces its contribution to catalysis (Figure 2b). While crystal contacts may contribute to this
129 displacement, size-exclusion chromatography coupled with small-angle X-ray scattering (SEC-
130 SAXS) confirmed that the designed lid adopts a similar conformation in solution (Supplementary
131 Figure 7).

132 Despite crystallization with 6NBT, no electron density corresponding to the transition-state
133 analogue was observed in the active site; instead, the density was best interpreted as a bound
134 glycerol molecule from the cryoprotectant. This suggests weak binding of the transition-state
135 analogue under crystallization conditions (Methods). Analysis of the active site revealed that the
136 catalytic base D153 adopts a similar rotameric configuration in both the crystal structure and
137 design model, whereas N46 adopts a rotameric state incompatible with transition-state
138 stabilization (Figure 4b). This aligns with the observation that mutation of N46 to Ala in KempTIM4-
139 2 causes only a small reduction in activity, while mutation of D153 abolishes activity (Figure 2b).
140 Additionally, the active site is poorly preorganized, with several amino acid side chains deviating
141 from the conformations designed to stabilize the transition state. These structural discrepancies
142 likely contribute to the modest catalytic efficiency of KempTIM4-2 and provide starting points for
143 future optimization.
144
145 **Structural Dynamics and Active-Site Accessibility Underpin Catalytic Activity**
146 To explore why only KempTIM4 and KempTIM4-2 were catalytically active, microsecond-
147 timescale molecular dynamics simulations were performed on these designs and the inactive
148 KempTIM9, both in the presence and absence of the transition-state analogue. KempTIM9 was
149 selected as a comparison due to its good expression (Supplementary Table 2), well-folded
150 structure and high melting temperature (67°C, Supplementary Figure 8), which rule out misfolding
151 as the cause of inactivity. The simulations revealed that the helical component of the lids, which
152 anchor the catalytic Asn, were rigid (Figure 5a). This rigidity stabilizes the transition-state
153 analogue within the binding pocket, maintaining catalytic interactions with the catalytic base for
154 much of the simulation, with active variants outperforming the inactive design KempTIM9 (Figure
155 5b). However, only KempTIM4 consistently maintained catalytic interactions with the H-bond
156 donor. Both KempTIM4-2 and KempTIM9 showed weaker contacts with this Asn residue, with

157 KempTIM4-2 failing to form this interaction throughout the simulation. This result aligns with the
158 lid displacement and unproductive rotameric state of N46 observed in the crystal structure, as
159 well as the small impact on activity when N46 was mutated to Ala in KempTIM4-2 (Figure 2d).

160 A key difference between the active designs and the inactive KempTIM9 was the
161 significantly narrower entrance to the active site in KempTIM9 (Figure 5b). Previous studies have
162 demonstrated that widening the active-site entrance enhances Kemp eliminase activity by
163 lowering the energy barriers associated with substrate entry and product release[24]. Thus, we
164 postulate that the inactivity of KempTIM9 is due to an insufficiently open active site, which likely
165 impedes substrate binding. By contrast, the lids of KempTIM4 and KempTIM4-2, which
166 incorporate five fragments rather than the single fragment found in KempTIM9 (Supplementary
167 Figure 3), may facilitate substrate entry. The additional inserted loops in these designs likely
168 contribute to active-site accessibility by providing greater flexibility to the lid in the unbound state
169 (Figure 5a), allowing the active site to open more readily. While other factors, such as incorrect
170 positioning of catalytic residues and lid misorientation, cannot be excluded, these findings suggest
171 that reducing barriers to substrate entry during lid generation could improve catalytic performance
172 in future designs.

173

174 **Opportunities for CANVAS design**

175 Natural TIM barrels use lids to bind substrates and form an active site conducive to
176 catalysis[25]. By contrast, minimal *de novo* TIM barrels lack such lids and are inherently
177 functionless. Using our CANVAS approach, we successfully designed a custom lid to form an
178 active site for a target reaction, resulting in the first enzyme constructed from a *de novo* TIM
179 barrel. While KempTIM4 exhibits modest activity, it matches the performance of many first-round
180 designs generated by traditional methods[17,19], despite the added complexity of engineering a
181 tailored lid comprising multiple loops to create an active site cavity. Further activity improvements

182 could be achieved by using the crystal structure reported here as a template for ensemble-based
183 design[24], a method that leverages crystallographic ensembles to design highly preorganized
184 active sites, leading to catalytic efficiency enhancements of several orders of magnitude[18]. Unlike
185 traditional enzyme design methods that rely on pre-existing scaffolds, CANVAS offers the
186 flexibility to customize lids for specific reaction requirements. It can also be readily adapted to
187 scaffold more than two catalytic residues, a task that becomes increasingly challenging with pre-
188 existing scaffolds as the number of catalytic groups grows. With CANVAS, we anticipate that any
189 minimal TIM barrel, including AI-generated virtual ones, could serve as a blank canvas for the
190 creation of a large diversity of *de novo* enzymes, unlocking the full potential of this versatile fold
191 for catalysis of a wide range of reactions.
192
193 **Methods**
194
195 ***Reagents and solutions.*** All experiments were conducted using analytical-grade chemicals, with
196 solutions prepared using double-distilled water.
197
198 ***Circular permutation of DeNovoTIM6-SB with inpainting.*** Circular permutation of
199 DeNovoTIM6-SB (PDB ID: 7OSV) was performed using the inpainting method described by Wang
200 et al.[21] The first residue of either the first or third β-strand was selected as the new N-terminus,
201 with the corresponding βα-loop deleted (Supplementary Figure 2). A new connection for the
202 corresponding βα-loop was modeled by inpainting 5 to 10 residues. Unresolved residues in the
203 crystal structure were remodeled, keeping intact the positions and identities of the remaining
204 residues. For each new connection, 100 designs were modeled, followed by relaxation and
205 scoring using the Rosetta protein design software[26]. Structures of each design were predicted
206 using AlphaFold2[16] with model 4 weights. Designs with a pLDDT value above 90 and a Rosetta

207 score below 3 Rosetta energy units per residue were kept. One design from each connection was
208 selected for experimental characterization (Supplementary Table 1). The proteins were named
209 NT6-CP1 (N-terminus on the first β-strand) and NT6-CP2 (N-terminus on the third β-strand).
210
211 ***Computational Enzyme Design.*** All calculations were performed with the Triad protein design
212 software (Protabit, Pasadena, CA, USA). Rotamer optimization was performed using a Monte
213 Carlo with simulated annealing search algorithm. Input structures were prepared for Triad
214 calculations via the *addH.py* application within Triad. The Kemp elimination transition state (TS)
215 structure was built using parameters from Privett *et al.*[17]. To provide sidechain conformations, a
216 backbone-independent conformer library (bda-bbind_1.0)[13] was used for theozyme placement,
217 and a backbone-independent rotamer library (bbind02.May.e2)[27] with expansions of ±1 standard
218 deviation around χ_1 and χ_2 was used for active-site repacking. Energies were calculated using a
219 modified version of the Phoenix energy function[12] consisting of a Lennard-Jones 12–6 van der
220 Waals term from the Dreiding II force field[28] with atomic radii scaled by 0.9, a direction-dependent
221 hydrogen bond term with a well depth of 8.0 kcal mol^{-1} and an equilibrium donor–acceptor
222 distance of 2.8 Å,[29] and an electrostatic energy term modeled using Coulomb's law with a
223 distance-dependent dielectric of 10. An energy benefit of −100 kcal mol^{-1} was applied when TS-
224 side-chain interactions satisfied catalytic contact geometries (Supplementary Table 5), as
225 described by Lassila *et al.*[13]
226
227 ***Theozyme placement.*** AlphaFold2 predictions of NT6-CP1 and NT6-CP2, along with the 7MCD
228 crystal structure, were used as backbone templates for theozyme placement. Amino-acid
229 residues with a C_α–C_β vector oriented towards the β-barrel interior and located at the C-terminus
230 of each β-strand were selected as positions for introduction of the catalytic base (Asp).
231 Neighbouring residues were mutated to alanine to avoid steric clashes with the TS

232 (Supplementary Table 6). TS poses were built using the contact geometries listed in
233 Supplementary Table 7, and those where the isoxazolic oxygen pointed toward the catalytic face
234 of the TIM barrel and remained coplanar with the catalytic base were selected (Supplementary
235 Figure 1). In parallel, the catalytic H-bond donor (Asn or Gln) was placed on a Gly-Asn/Gln-Gly
236 tripeptide scaffold, and TS poses were built using H-bond donor side-chain contacts specified in
237 Supplementary Table 7. The resulting TS was superimposed with the TS from the catalytic base
238 calculation to create the full theozyme, positioning the catalytic base within the minimal TIM barrel
239 and the H-bond donor in the empty space above the catalytic face. The Gly residues from the
240 tripeptide were then deleted, preserving only the Asn/Gln residue, whose alpha carbon atom
241 served as an anchor for the designed lid.
242
243 ***Lid design.*** For all theozyme placements, peptide connections between the catalytic H-bond
244 donor residue and the minimal TIM barrel structures were designed using RFdiffusion[14]. For each
245 position of the H-bond donor alpha carbon, various βα-loops were selected as insertion points for
246 a primary peptide fragment, and multiple fragment lengths were sampled to connect to the
247 catalytic residue while preserving the TIM-barrel topology. Additional βα-loops were sometimes
248 elongated to improve interactions with the primary fragment (Supplementary Figure 3). For each
249 inserted primary fragment, 100–200 structures were sampled. Primary fragments lacking
250 secondary structure, causing steric clashes with the TS, or displacing the H-bond donor from its
251 theozyme position, were discarded (Supplementary Figure 1). The filtered fragments were
252 sequence-optimized using ProteinMPNN[15] using a temperature factor of 0.1, ensuring the minimal
253 TIM barrel template sequence remained unchanged. Cysteine and methionine residues were
254 excluded during sequence design. Structures of designed sequences were predicted using
255 AlphaFold2[16] or ColabFold v1.3.0[30] with all five model weights, and those displaying an average

256 pLDDT value >90 and backbone RMSD < 3 Å relative to the RFdiffusion model were selected for
257 active-site repacking.
258
259 ***Active-site repacking.*** AlphaFold2 models generated above were used for a theozyme
260 placement step, as described previously, to verify that contacts between catalytic residues and
261 the TS could be formed on these scaffolds. Active-site repacking calculations were then
262 performed on the structures with this new theozyme placement. In these calculations, the TS was
263 translated by ±0.4 Å along each Cartesian coordinate in 0.2-Å steps and rotated about all three
264 axes (origin at the TS geometric center) in 5° increments over a 10° range (clockwise and
265 counterclockwise). This resulted in a total combinatorial search space of 15,625 possible poses.
266 Residues near the catalytic amino acids and TS were designated as design positions. During
267 repacking, these positions were allowed to sample rotamers of various amino acids, favoring
268 hydrophobic and aromatic residues, while the identities of the catalytic residues remained fixed
269 (Supplementary Table 8).
270
271 ***Computational library design***. After repacking, the CLEARSS computational library design
272 algorithm[31] was used to generate a combinatorial library comprising the most favorable amino
273 acids predicted by Triad at each designed active-site position. In this method, libraries of a specific
274 size configuration are generated from a pre-scored list of sequences using the highest probability
275 set of amino acids at each position based on the sum of their Boltzmann weights. Using the list
276 of energy-ranked sequences from active-site repacking as input, libraries of 192 sequences were
277 generated. Rotamer configurations for each sequence in the library were optimized using the
278 *cleanSequences.py* application within Triad to find the lowest-energy conformation of each
279 sequence on its respective backbone, generating “cleaned” structures. To compare energies, the
280 energy difference between each “cleaned” structure and a corresponding all-Gly structure in the

absence of the TS was calculated. These energies are reported in Supplementary Table 9. Structures were cleaned with and without the TS to evaluate preorganization, as described previously[18].

Design filtering. Energies of the TS-bound and unbound "cleaned" structures were calculated using Triad. Dihedral angles of catalytic residues relative to the TS, as well as the solvent-accessible surface area (SASA) of the TS, were analyzed using PyMOL (v2.3.0, Schrödinger, LLC). Bottleneck radii of active-site entrances were evaluated using CAVER v3.0[32], and the number of residues mutated to glycine or methionine was counted. Designs were filtered based on the following criteria (Supplementary Table 9): energy difference between bound and unbound structures below 0 kcal mol^{-1}, catalytic residue dihedral angles outside 50° to 130° and –50° to –130°, SASA of TS between 80 and 200 $Å^2$, < 6 Met and < 4 Gly, bottleneck radius of active-site entrance > 0.9 Å, and < 7 non-preorganized residues. Designs passing all criteria were used for structure prediction with AlphaFold2 or ColabFold (v1.3.0), where designs with a backbone RMSD < 3 Å relative to the structure cleaned by Triad, overall pLDDT > 90, and lid pLDDT > 85 were chosen for experimental characterization (Supplementary Table 1).

KempTIM4 redesign. Sequence optimization of KempTIM4 was performed using ProteinMPNN with LigandMPNN weights[33]. Amino-acid identities of all residues designed during active-site repacking of KempTIM4 were retained (Supplementary Table 8). Four sequences were generated, omitting cysteine. The generated sequences were used for structure prediction with ColabFold v1.3.0[30] using all five model weights. Predictions were filtered based on their all-atom RMSD to the input KempTIM4 AlphaFold2 model. Theozyme placement and active-site repacking for each sequence were performed as described earlier using the backbones of AlphaFold2

305 models that passed filtering. Following energy calculations with Triad, a single design,
306 KempTIM4-2, was selected for experimental characterization (Supplementary Table 1).
307
308 ***Cloning and generation of constructs.*** Genes for all proteins were ordered as codon-optimized
309 fragments flanked by restriction sites for *Nde*I and *Xho*I from Twist Bioscience (South San
310 Francisco, CA, USA). After digestion with *Nde*I and *Xho*I, the gene fragment was ligated into
311 pET21b(+) for KempTIM1–5, or pET29b(+) for KempTIM6–9. Catalytic knockout mutations were
312 introduced by a modified QuikChange PCR protocol utilizing KAPA Polymerase followed by *Dpn*I
313 digestion. *E. coli* Top10 or BL21 (DE3) cells were transformed with ligated vector or reaction
314 mixture and plated on Lysogeny Broth (LB) agar plates containing 100 µg mL^{-1} ampicillin or 50
315 µg mL^{-1} kanamycin as selection markers. Single colonies were used to inoculate 5 mL LB
316 supplemented with 100 µg mL^{-1} ampicillin or 50 µg mL^{-1} kanamycin. After overnight growth
317 (37 °C, 250 rpm), cells were harvested by centrifugation and DNA was isolated using NucleoSpin
318 Plasmid EasyPure-Kit (Machery & Nagel) according to the manufacturer's protocol. Vector
319 assembly and introduction of mutations were validated by sequencing (Eurofins Genomics) using
320 standard T7 primers.
321
322 ***Protein expression and purification.*** *E.coli* BL21 (DE3) cells (Novagen) were transformed with
323 plasmid, plated on agar plates containing antibiotic, and incubated overnight at 37 °C. Single
324 colonies were picked to inoculate precultures and incubated at 30–37 °C overnight. On the next
325 day, 1 L LB supplemented with antibiotic was inoculated with 10 mL of the preculture and
326 incubated at 37 °C until OD600 reached a value between 0.6 and 0.8. Overexpression was
327 induced by adding isopropyl-β-thiogalactoside (IPTG) to a final concentration of 0.1 mM for
328 KempTIM1–5 or 1 mM for KempTIM6–9 followed by overnight incubation at 16–20 °C with

329 shaking. Cells were harvested by centrifugation and pellets were either frozen at –20 °C or used
330 directly for purification.

331 Two different purification schemes were employed. In the first scheme, cell pellets were
332 resuspended in 35 mL buffer A (50 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 mM
333 imidazole). Resuspended cells were lysed by sonication (Branson Ultrasonic Sonifier 250, output
334 4, duty cycle 40%, 3 × 3 min) and centrifuged. The supernatant was loaded onto a HisTrapHP
335 column (5 mL, Cytiva Life Science) equilibrated with buffer A and coupled to an ÄKTApure system
336 (Cytiva Life Science). After washing with 10 column volumes (CV) of buffer A, the protein was
337 eluted with a linear gradient over 20 CV to 60 % buffer B (50 mM sodium phosphate pH 7.0, 100
338 mM NaCl, 500 mM imidazole). For KempTIM4-2L, which contains an additional GSG tripeptide
339 linker between its His-tag and first TIM barrel residue (Supplementary Table 1), an additional
340 TEV-cleavage over night was performed during a dialysis against buffer C (50 mM of sodium
341 phosphate pH 7.0, 100 mM NaCl). The dialyzed protein was loaded onto a HisTrapHP column
342 (5 mL, Cytiva Life Science) equilibrated with buffer C and coupled to an ÄKTApure system (Cytiva
343 Life Science) and the flowthrough was collected. Fractions containing the protein were pooled,
344 concentrated with a centrifugal concentrator, and loaded onto a HiLoad 26/600 Superdex 75
345 preparative grade column (Cytiva Life Sciences) preequilibrated in buffer C and eluted with 1 CV
346 buffer C.

347 In the second scheme, cell pellets were resuspended in 8 mL buffer D (5 mM imidazole in
348 100 mM potassium phosphate buffer at pH 8.0) supplemented with 1 mg mL^{-1} lyophilized
349 lysozyme (MP Biomedicals) and 1 U mL^{-1} benzonase nuclease (Merck Millipore). Cells were
350 homogenized using an Avestin EmulsiFlex-B15 cell disruptor. Proteins were purified from the
351 lysate by immobilized metal affinity chromatography using Ni-NTA agarose (Qiagen) pre-
352 equilibrated with buffer D in individual Econo-Pac gravity-flow columns (Bio-Rad). Contaminants
353 were washed away using buffer E (10 mM imidazole in 100 mM potassium phosphate buffer pH

354 8.0) and then 20 mM imidazole in the same buffer. Proteins were eluted with 5 mL of buffer F
355 (250 mM imidazole in 100 mM potassium phosphate buffer pH 8.0). Proteins were further
356 subjected to gel filtration in buffer G (25 mM HEPES pH 7.25 and 100 mM sodium chloride) using
357 an ENrich SEC 650 size-exclusion chromatography column (Bio-Rad). Purification was checked
358 by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and fractions
359 containing the protein of interest were pooled. Protein concentration was determined
360 spectrophotometrically using the absorption at 280 nm and applying Beer-Lambert's law using
361 calculated extinction coefficients obtained from the ExPAsy ProtParam tool
362 (https://web.expasy.org/protparam/).
363
364 ***Far-UV circular dichroism (CD) spectroscopy.*** CD spectra were collected with a Jasco J-710
365 or J-815 spectrometer. Spectra were recorded using a 1-mm quartz cuvette at 20 °C with
366 scanning speed 10–100 nm min^{-1}, bandwidth 1 nm, response time 1 s. Experiments were
367 performed in 20 mM sodium phosphate buffer pH 7.0 or 50 mM sodium phosphate buffer pH 8.0
368 supplemented with 100 mM sodium chloride, using a protein concentration of approximately
369 0.2 mg mL^{-1}. For each protein, 3 to 10 spectra were accumulated and averaged. For data
370 normalization, a buffer spectrum was subtracted and the signal was converted to mean residue
371 molar ellipticity using $[\theta_{MRE}] = \theta/(n \times d \times c)$, where n is the number of residues in the protein, θ the
372 collected ellipticity in mdeg, d the path length in mm, and c the protein concentration in M. Thermal
373 denaturation assays were performed by heating the samples from 20 to 95 °C at a rate of 1 °C
374 per minute and ellipticity at 222 nm was measured every 1 °C. Melting temperature (T_m) was
375 determined following the protocol by Greenfield[34].
376
377 ***Size-exclusion chromatography-multi angle light scattering.*** SEC-MALS measurements
378 were performed as described in Beck *et al.*[8] Each protein was measured at 2 mg mL^{-1} except

379 KempTIM4 N46A (1 mg mL^{-1}). For all measurements, buffer C and an injection volume of 50 µL
380 were used.
381
382 ***Steady-state kinetics.*** All assays were performed in buffer C or G at 27 °C in a Spark (Tecan)
383 or Synergy H1 (Biotek) plate reader. Enzyme concentrations varied from 5 to 80 µM. Reactions
384 (200 µL final volume) were initiated by addition of varying concentrations (0.03–2 mM) of
385 5-nitrobenzisoxazole (abcr or AAblocks) dissolved in methanol (final methanol concentration
386 10 %). Product formation was monitored at 380 nm (ε = 15,800 M^{-1} cm^{-1})[17] in individual wells of a
387 96-well plate (Nunc or Greiner Bio-One). Path lengths for each well were calculated ratiometrically
388 using the difference in absorbance at 900 and 975 nm. Control measurements without protein
389 were conducted for each substrate concentration and subtracted from the enzyme
390 measurements. Linear phases of the kinetic traces were used to measure initial reaction rates. If
391 the enzyme showed saturation within the substrate concentration used, the data were fitted to the
392 Michealis-Menten model using ($v_0 = (v_{max} \times [S]) / (K_M + [S])$ where v_0 is the initial velocity rate, v_{max}
393 is the maximal velocity rate, [S] is the substrate concentration, K_M is the Michaelis constant. The
394 catalytic constant k_{cat} was calculated with $k_{cat} = v_{max} / [E]$ where [E] is the enzyme concentration.
395 If saturation was not achieved within the substrate's solubility limit, a linear equation ($v_0 = k_{cat}/K_M$
396 $\times$ [S]) was used.
397
398 ***Inhibition assays.*** Inhibition assays of KempTIM4-2 were performed in 25 mM HEPES buffer
399 (pH 7.25) with 100 mM NaCl and 1 % DMSO at 27 °C using a Biotek Synergy H1 multimode plate
400 reader. Reactions (200 µL final volume) were initiated by addition varying concentrations (0.25–
401 2 mM) of 5-nitrobenzisoxazole (AAblocks) dissolved in methanol (10 % final concentration) to the
402 enzyme mixed with varying concentrations (0.5–1 mM) of 6-nitro-1H-benzotriazole (Aaron
403 Chemicals) in wells of a 96-well plate (Greiner Bio-One). The formation of the product was

404 monitored spectrophotometrically at 380 nm (ε = 15,800 M^{-1} cm^{-1}). Path lengths for each well
405 were calculated ratiometrically using the difference in absorbance at 900 and 975 nm. Linear
406 phases of kinetic traces were used to measure initial reaction rates. Michaelis-Menten and
407 Lineweaver-Burk plots were created from the resulting data to determine the type of inhibition.
408
409 ***Crystallization and structure determination.*** KempTIM4-2L (Supplementary Table 1) was
410 purified using the first purification scheme, with the final size-exclusion chromatography step
411 performed in 20 mM HEPES buffer pH 8.0 supplemented with 20 mM NaCl. Crystallization
412 screens were set up using the sitting-drop vapor diffusion method with a Phoenix pipetting robot
413 (Art Robbins Instruments) and commercially available sparse-matrix screens (NeXtal) in 96-well
414 sitting-drop plates (3-drop Intelli-Plates, Art Robbins Instruments). Protein and reservoir solutions
415 were mixed in ratios of 1:1, 2:1, and 1:2. The protein concentration was 12 mg mL^{-1},
416 supplemented with 5 mM 6NBT (5% DMSO). Crystals were grown at 293 K, and diffracting
417 crystals were obtained after 40 days in 4 M sodium formate. Cryoprotection was achieved by
418 adding glycerol to a final concentration of 25%, and crystals were mounted using cryo-loops on
419 SPINE standard bases and flash-cooled in liquid nitrogen. Diffraction data were collected at 100
420 K at beamline BL 14.1 with a Pilatus3 S 6M detector at the BESSY II synchrotron (Helmholtz-
421 Zentrum Berlin). Data processing was performed using X-ray Detector Software APP3
422 (XDSAPP3)[35] with XDS[36], and data quality was assessed using phenix.xtriage[37]. Phases were
423 solved by molecular replacement using the AlphaFold2 prediction as search model with Phaser[38].
424 The resulting model was manually rebuilt using Coot[39] and refined with phenix.refine[37] in an
425 iterative process. Refinement statistics and crystallographic data are shown in Supplementary
426 Table 4. Coordinates and structure factors were validated and deposited in the PDB under
427 accession code 9QKX.
428

429 ***Size-Exclusion Chromatography Small-angle X-ray Scattering (SEC-SAXS).*** SEC-SAXS
430 measurements were conducted at the BioSAXS beamline BM29 of the ESRF in Grenoble,
431 France. A Superdex 75 Increase 10/300 GL column (Cytiva Life Sciences) was used with a flow
432 rate of 0.8 mL min^{-1} in buffer C. KempTIM4-2L (Supplementary Table 1) was measured at a
433 concentration of 5 mg mL^{-1}, with an injection volume of 100 µL. Data processing and analysis
434 were performed using ATSAS 3.2.1[40,41]. For analysis, AlphaFold2 models of KempTIM4-2 and
435 NT6-CP2, as well as an AlphaFlow[42] model of KempTIM4-2 with a displaced lid obtained using
436 the base molecular dynamics weights were used. For each of these models a theoretical
437 scattering curve were generated and fitted to the experimental data using CRYSOL[43].
438
439 ***Molecular dynamics (MD) simulations.*** The Amber 2020 software (http://ambermd.org/) with
440 the ff19SB protein force field[44], gaff2 ligand force field[45], and OPC water force field[46] was used for
441 all simulations. A cutoff of 10 Å was applied to electrostatics modeled using the particle mesh
442 Ewald method[47]. All MD trajectories were run with a time step of 2 fs. Models of each design
443 generated using Triad as described above were prepared for MD. MD trajectories were run for
444 each enzyme in the presence and absence of bound transition-state analogue 6-
445 nitrobenzotriazole (6NBT). Parameters for 6NBT were generated using the Antechamber
446 package[48]. Hydrogen atoms were added using Reduce[49], and the prepared structures were
447 solvated with OPC water in a truncated octahedral box with periodic boundary conditions where
448 the distance between the protein surface and the box edges was set to 10 Å. The addions2
449 algorithm in Amber was used to place counterions to neutralize the system. The system was first
450 minimized through the method of steepest descent using a force constant of 500 kcal mol^{-1} $Å^{-2}$.
451 The system was then heated to a temperature of 300 K over 240 ps while restraints were gradually
452 removed, followed by equilibration under an NPT ensemble at 300 K and 1 bar for 10 ns and
453 subsequent equilibration under an NVT ensemble at 300 K for another 10 ns. Constant

454 temperature and pressure were achieved using the Langevin thermostat[50] and Berendsen
455 barostat[51], respectively. Following equilibration, triplicate 1-μs production simulations were run.
456 For analyzing bottleneck radii, 1000 snapshots separated by 1 ns each were extracted from the
457 production trajectory. A minimum cutoff of 0.9 Å was used to identify active-site entrance
458 bottlenecks in the unbound MD trajectories using CAVER 3.0[32]. To analyze catalytic interactions
459 and RMSF, 5×10^4 snapshots separated by 20 ps each were extracted from the production
460 trajectory. Catalytic interactions and RMSF values were extracted using CPPTRAJ and PYTRAJ,
461 respectively[52].
462

Figures

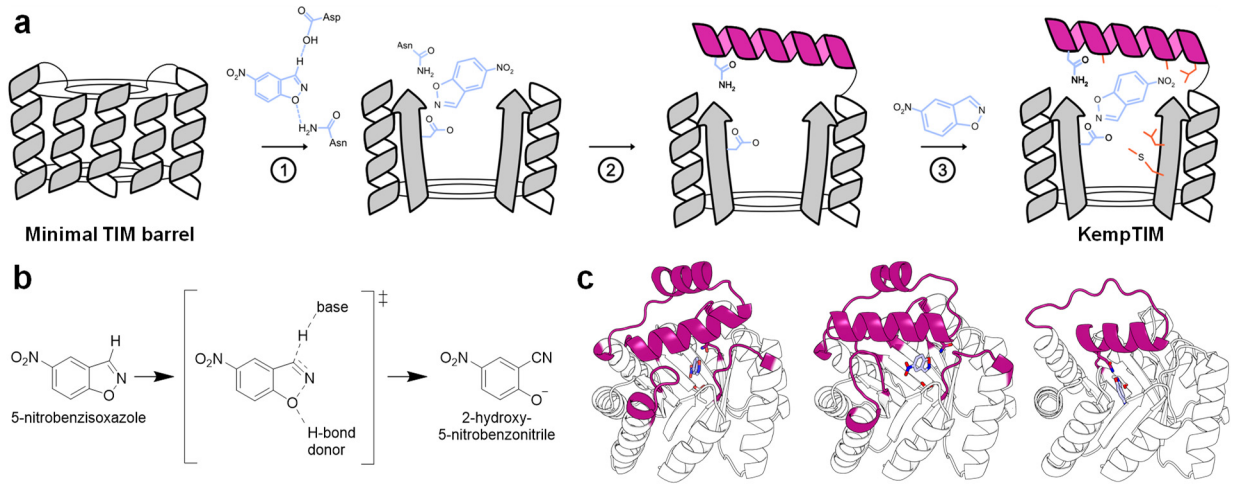

Figure 1. *De novo* enzyme design using CANVAS. (a) CANVAS follows three general steps: (1) placement of the theozyme (blue) onto the minimal TIM-barrel scaffold using a single catalytic residue as an anchor; (2) construction of a custom lid (magenta) to form the active-site pocket and anchor the second 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.

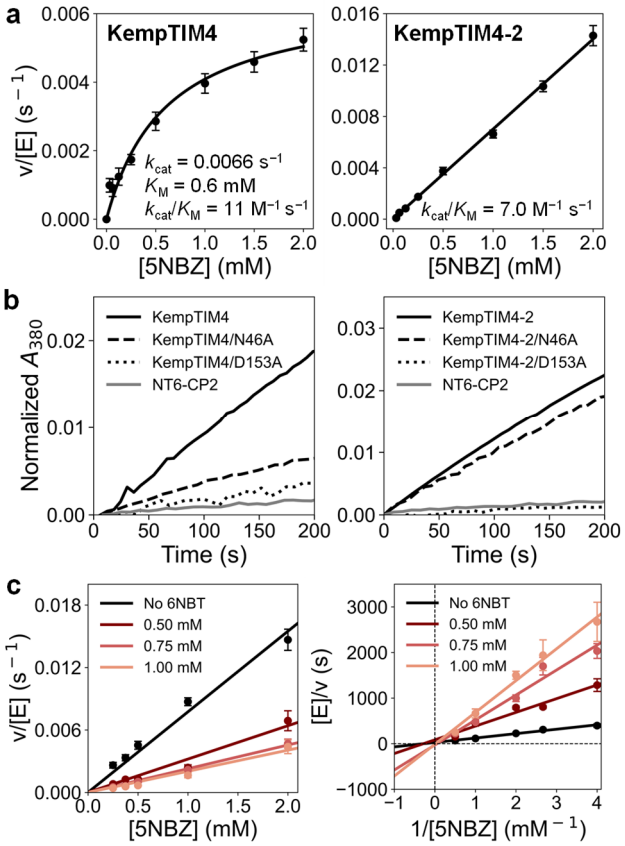

Figure 2. Kinetic characterization of KempTIMs. (a) Michaelis–Menten plots of normalized initial rates as a function of 5-nitrobenzisoxazole (5-NBZ) concentrations. Data represent the average of 18 or nine replicate measurements from five or four independent protein batches for KempTIM4 and KempTIM4-2, respectively (mean ± SEM). (b) Reaction progress curves ([5NBZ] = 2 mM) demonstrate the contributions of designed catalytic residues D153 and N46 to catalysis. (c) Michaelis-Menten (left) and Lineweaver-Burk (right) plots reveal that the transition-state analogue 6NBT acts as a competitive inhibitor of KempTIM4-2. Data represent the average of four replicate measurements from a single protein batch (mean ± SEM).

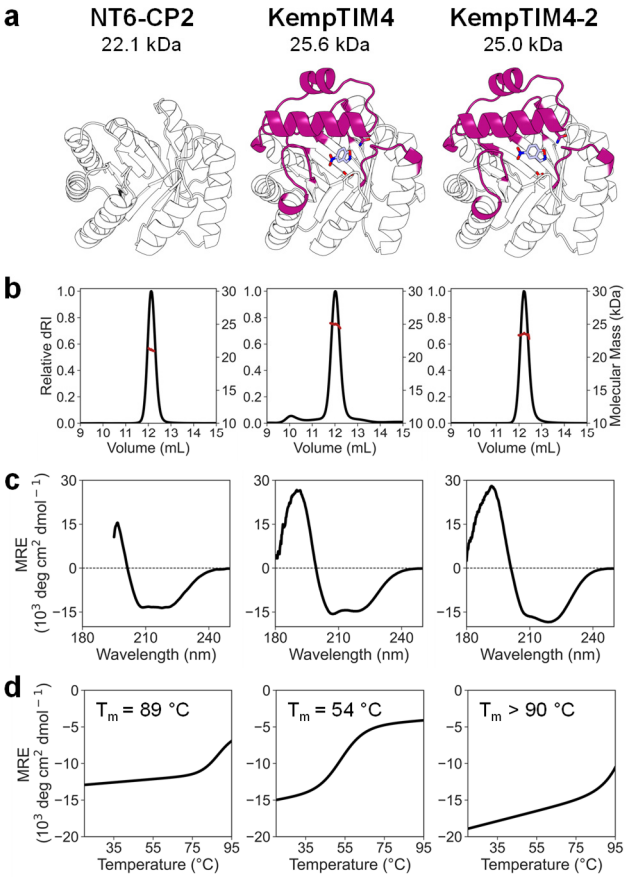

Figure 3. Structural characterization. (a) Computational models of minimal TIM-barrel NT6-CP2 and active KempTIMs with the designed lid shown in magenta and the theozyme as sticks. (b) SEC-MALS indicates that the proteins are predominantly monomeric, with molecular weights close to their expected values. (c) CD spectra reveal a mixed αβ signal characteristic of TIM barrels. MRE: mean residue ellipticity. (d) Melting curves demonstrate that KempTIM4 is substantially destabilized compared to NT6-CP2. Redesign to yield KempTIM4-2 restores stability.

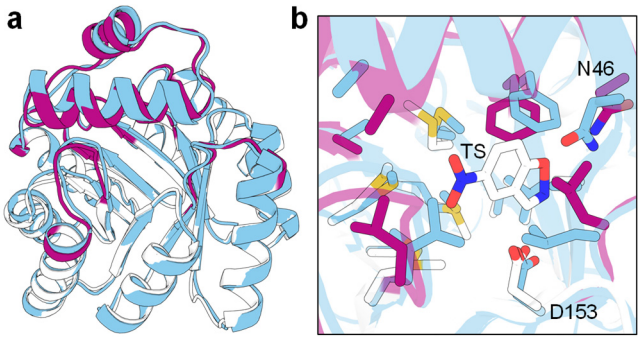

Figure 4. Crystal structure of KempTIM4-2. (a) The crystal structure (blue) aligns closely with the design model (minimal TIM barrel and lid colored white and magenta, respectively). The backbone RMSD is 0.86 Å for the full structure and 1.34 Å for the lid. (b) Active-site residues in the crystal structure (blue) and design model (white and magenta) show greater divergence, possibly due to the absence of a bound transition-state analogue in the crystal structure. While the catalytic base D153 adopts a similar rotamer in both structures (side chain RMSD = 0.76 Å), the catalytic H-bond donor N46 adopts a rotameric configuration incompatible with transition-state (TS) stabilization (side chain RMSD = 3.09 Å).

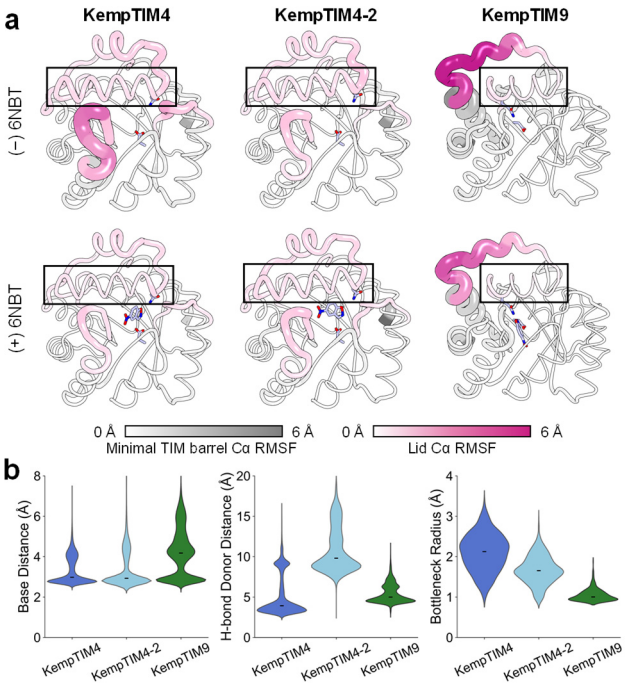

Figure 5. Molecular dynamics in the presence and absence of transition-state analogue 6NBT. (a) Root-mean-square fluctuations (RMSF) of the protein backbone, represented as putty plots, reveal that the lid α-helix anchoring the catalytic H-bond donor (boxed) is rigid, whereas loops exhibit greater flexibility. 6NBT and catalytic residues are depicted as sticks. (b) Distributions of catalytic contact distances between 6NBT and catalytic base or H-bond donor are shown alongside the active-site entrance bottleneck radius, presented as violin plots, with the median value indicated by a black line.

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 and thank the beamline staff at BESSY 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

Birte Höcker: 0000-0002-8250-9462

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, doi:https://doi.org/10.1016/S0022-2836(02)00649-6 (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, doi:https://doi.org/10.1016/S0958-1669(00)00230-5 (2001).

3 Wierenga, R. K., Kapetaniou, E. G. & Venkatesan, R. Triosephosphate isomerase: a highly evolved biocatalyst. *Cellular and Molecular Life Sciences* **67**, 3961-3982, doi:10.1007/s00018-010-0473-9 (2010).

4 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, doi:10.1038/nchembio.1966 (2016).

5 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, doi:https://doi.org/10.1002/pro.4249 (2022).

6 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, doi:https://doi.org/10.1016/j.jmb.2021.167153 (2021).

7 Chu, A. E., Fernandez, D., Liu, J., Eguchi, R. R. & Huang, P.-S. De Novo Design of a Highly Stable Ovoid TIM Barrel: Unlocking Pocket Shape towards Functional Design. *BioDesign Research* **2022**, 9842315, doi:https://doi.org/10.34133/2022/9842315 (2022).

8 Beck, J., Shanmugaratnam, S. & Höcker, B. Diversifying de novo TIM barrels by hallucination. *Protein Science* **33**, e5001, doi:https://doi.org/10.1002/pro.5001 (2024).

9 Kordes, S., Beck, J., Shanmugaratnam, S., Flecks, M. & Höcker, B. Physics-based approach to extend a de novo TIM barrel with rationally designed helix-loop-helix motifs. *Protein Engineering, Design and Selection* **36**, gzad012, doi:10.1093/protein/gzad012 (2023).

10 Wiese, J. G., Shanmugaratnam, S. & Höcker, B. Extension of a de novo TIM barrel with a rationally designed secondary structure element. *Protein Science* **30**, 982-989, doi:https://doi.org/10.1002/pro.4064 (2021).

11 Tantillo, D. J., Jiangang, C. & Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Current Opinion in Chemical Biology* **2**, 743-750, doi:https://doi.org/10.1016/S1367-5931(98)80112-9 (1998).

12 Lee, F. S., Anderson, A. G. & Olafson, B. D. Benchmarking TriadAb using targets from the second antibody modeling assessment. *Protein Engineering, Design and Selection* **36**, gzad013, doi:10.1093/protein/gzad013 (2023).

13 Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 16710-16715, doi:10.1073/pnas.0607691103 (2006).

14 Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100, doi:10.1038/s41586-023-06415-8 (2023).

15 Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56, doi:10.1126/science.add2187 (2022).

16 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

17 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**, 3790-3795, doi:10.1073/pnas.1118082108 (2012).
18 Rakotoharisoa, R. V. *et al.* Design of Efficient Artificial Enzymes Using Crystallographically Enhanced Conformational Sampling. *J Am Chem Soc* **146**, 10001-10013, doi:10.1021/jacs.4c00677 (2024).
19 Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).
20 Anand, N. *et al.* Protein sequence design with a learned potential. *Nature Communications* **13**, 746, doi:10.1038/s41467-022-28313-9 (2022).
21 Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387-394, doi:10.1126/science.abn2100 (2022).
22 Zhang, S., Zhang, J., Luo, W., Wang, P. & Zhu, Y. A preorganization oriented computational method for de novo design of Kemp elimination enzymes. *Enzyme and Microbial Technology* **160**, 110093, doi:https://doi.org/10.1016/j.enzmictec.2022.110093 (2022).
23 Sumida, K. H. *et al.* Improving Protein Expression, Stability, and Function with ProteinMPNN. *Journal of the American Chemical Society* **146**, 2054-2061, doi:10.1021/jacs.3c10941 (2024).
24 Broom, A. *et al.* Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nature Communications* **11**, 4808, doi:10.1038/s41467-020-18619-x (2020).
25 Sterner, R. & Höcker, B. Catalytic Versatility, Stability, and Evolution of the (βα)8-Barrel Enzyme Fold. *Chemical Reviews* **105**, 4038-4055, doi:10.1021/cr030191z (2005).
26 Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**, 545-574, doi:10.1016/B978-0-12-381270-4.00019-6 (2011).
27 Dunbrack, R. L. Rotamer Libraries in the 21st Century. *Current Opinion in Structural Biology* **12**, 431-440, doi:https://doi.org/10.1016/S0959-440X(02)00344-5 (2002).
28 Mayo, S. L., Olafson, B. D. & Goddard, W. A. Dreiding - a Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* **94**, 8897-8909, doi:Doi 10.1021/J100389a010 (1990).
29 Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in protein design. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10172-10177 (1997).
30 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679-682, doi:10.1038/s41592-022-01488-1 (2022).
31 Allen, B. D., Nisthal, A. & Mayo, S. L. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19838-19843, doi:10.1073/pnas.1012985107 (2010).
32 Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Computational Biology* **8**, e1002708, doi:10.1371/journal.pcbi.1002708 (2012).
33 Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *bioRxiv*, 2023.2012.2022.573103, doi:10.1101/2023.12.22.573103 (2023).
34 Greenfield, N. J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nature Protocols* **1**, 2733-2741, doi:10.1038/nprot.2006.229 (2006).

35 Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. XDSAPP2.0. *Journal of applied crystallography* **49**, 1085-1092, doi:doi:10.1107/S1600576716004416 (2016).
36 Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125-132, doi:doi:10.1107/S0907444909047337 (2010).
37 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta crystallographica. Section D, Structural biology* **75**, 861-877, doi:10.1107/S2059798319011471 (2019).
38 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
39 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
40 Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of applied crystallography* **50**, 1545-1553, doi:doi:10.1107/S1600576717011438 (2017).
41 Manalastas-Cantos, K. *et al.* ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of applied crystallography* **54**, 343-355, doi:doi:10.1107/S1600576720013412 (2021).
42 Jing, B., Berger, B., Jaakkola, T. AlphaFold Meets Flow Matching for Generating Protein Ensembles. *arXiv* **2402**, 04845 (2024).
43 Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* **50**, 1212-1225, doi:doi:10.1107/S1600576717007786 (2017).
44 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **78**, 1950-1958, doi:https://doi.org/10.1002/prot.22711 (2010).
45 He, X., Man, V. H., Yang, W., Lee, T.-S. & Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *The Journal of Chemical Physics* **153**, 114502, doi:10.1063/5.0019056 (2020).
46 Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **5**, 3863-3871, doi:10.1021/jz501780a (2014).
47 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092, doi:10.1063/1.464397 (1993).
48 Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25**, 247-260, doi:https://doi.org/10.1016/j.jmgm.2005.12.005 (2006).
49 Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **285**, 1735-1747, doi:10.1006/jmbi.1998.2401 (1999).
50 Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics* **114**, 2090-2098, doi:10.1063/1.1332996 (2001).
51 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690, doi:10.1063/1.448118 (1984).
52 Roe, D. R. & Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095, doi:10.1021/ct400341p (2013).