

1   **Evolution of an enzyme conformational ensemble guides design of an efficient**  
2   **biocatalyst**

3

4   **Aron Broom,<sup>1,†</sup> Rojo V. Rakotoharisoa,<sup>1,†</sup> Michael C. Thompson,<sup>2</sup> Niayesh Zarifi,<sup>1</sup> Erin**  
5   **Nguyen,<sup>1</sup> Nurzhan Mukhametzhhanov,<sup>1</sup> Lin Liu,<sup>2</sup> James S. Fraser<sup>2</sup> & Roberto A. Chica<sup>1,\*</sup>**

6

7   <sup>1</sup> Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie,  
8   K1N 6N5, Ottawa, Ontario, Canada

9   <sup>2</sup> Department of Bioengineering and Therapeutic Science, University of California, San  
10 Francisco, San Francisco, California 94158, United States

11

12   <sup>†</sup> These authors contributed equally to this work.

13   <sup>\*</sup> To whom correspondence should be addressed:

14   Roberto A. Chica, Email: [rchica@uottawa.ca](mailto:rchica@uottawa.ca)

15

16    **Abstract**

17    The creation of artificial enzymes is a key objective of computational protein design. Although *de*  
18    *novo* enzymes have been successfully designed, these exhibit low catalytic efficiencies, requiring  
19    directed evolution to improve activity. Here, we used room-temperature X-ray crystallography to  
20    study changes in the conformational ensemble during evolution of the designed Kemp eliminase  
21    HG3 ( $k_{\text{cat}}/K_M$  160 M<sup>-1</sup>s<sup>-1</sup>). We observed that catalytic residues were increasingly rigidified, the  
22    active site became better pre-organized, and its entrance was widened. Based on these  
23    observations, we engineered HG4, an efficient biocatalyst ( $k_{\text{cat}}/K_M$  120,000 M<sup>-1</sup>s<sup>-1</sup>) containing  
24    active-site mutations found during evolution but not distal ones. HG4 structures revealed that its  
25    active site was pre-organized and rigidified for efficient catalysis. Our results show how directed  
26    evolution circumvents challenges inherent to enzyme design by shifting conformational ensembles  
27    to favor catalytically-productive sub-states, and suggest improvements to the design methodology  
28    that incorporate ensemble modeling of crystallographic data.

29     **Introduction**

30           Enzymes are the most efficient catalysts known, accelerating chemical reactions by up to  
31       26 orders of magnitude<sup>1</sup> while displaying unmatched selectivity. The ability to create, from  
32       scratch, an efficient artificial enzyme for any desired chemical reaction (i.e., a *de novo* enzyme) is  
33       a key objective of computational protein design. Progress towards this goal has been made over  
34       the past few decades following the development of computational enzyme design algorithms.<sup>2,3</sup>  
35       These methods have been used to create *de novo* enzymes for a variety of model organic  
36       transformations including the Kemp elimination,<sup>4,5</sup> retro-aldol,<sup>6,7</sup> Diels-Alder,<sup>8</sup> ester hydrolysis,<sup>9</sup>  
37       and Morita-Baylis-Hilman<sup>10</sup> reactions. Although successful, catalytic activities of *de novo*  
38       enzymes have been modest, with  $k_{\text{cat}}/K_M$  values being several orders of magnitude lower than those  
39       of natural enzymes.<sup>11,12</sup> In addition, structural analyses of designed enzymes have revealed  
40       important deficiencies in the computational methodologies, resulting in inaccurate predictions of  
41       catalytic and ligand-binding interactions,<sup>5</sup> and thereby low success rates,<sup>4,6,8</sup> emphasizing the need  
42       for continued development of robust enzyme design algorithms.

43           To improve the catalytic activity of designed enzymes, researchers have used directed  
44       evolution. This process has yielded artificial enzymes displaying catalytic efficiencies approaching  
45       those of their natural counterparts, and provided valuable information about the structural  
46       determinants of efficient catalysis.<sup>4,13-15</sup> During evolution, active-site residues, including designed  
47       catalytic amino acids, were often mutated, leading to enhanced catalysis via optimization of  
48       catalytic contacts, ligand binding modes, and transition-state complementarity.<sup>13-15</sup> Directed  
49       evolution has also yielded beneficial mutations at positions remote from the active site. Distal  
50       mutations have been shown to enhance catalysis by shifting the populations of conformational sub-  
51       states that enzymes sample on their energy landscape towards those that are more catalytically

52 active.<sup>16-18</sup> Therefore, a better understanding of enzyme conformational ensembles, including the  
53 effect of mutations on the population of sub-states, could provide valuable insights to aid in the  
54 development of robust computational enzyme design methodologies.

55 Here, we study changes in the conformational ensemble along the evolutionary trajectory  
56 of the *de novo* Kemp eliminase HG3 ( $k_{\text{cat}}/K_M$  160 M<sup>-1</sup>s<sup>-1</sup>) using room-temperature X-ray  
57 crystallography. We observe that during evolution, catalytic residues were increasingly rigidified  
58 through improved packing, the active site became better pre-organized to favor productive binding  
59 of the substrate, and the active-site entrance was widened to facilitate substrate entry and product  
60 release. Based on these observations, we generated a variant that contains all mutations necessary  
61 to establish these structural features, which are found at positions within or close to the active site.  
62 This variant, HG4, is >700-fold more active than HG3, with a catalytic efficiency on par with that  
63 of the average natural enzyme ( $k_{\text{cat}}/K_M$  120,000 M<sup>-1</sup>s<sup>-1</sup>). Crystallographic analysis of HG4 reveals  
64 that mutations proximal to the active site are sufficient to alter the conformational ensemble for  
65 enrichment of catalytically-competent sub-states. Lastly, we demonstrate that HG4 can be  
66 successfully designed using a crystallographically-derived ensemble of backbone templates  
67 approximating conformational flexibility, but not with the single template used to design HG3,  
68 offering insights for improving enzyme design methodologies.

69

## 70 **Results**

### 71 *HG series of Kemp eliminases*

72 Perhaps the most successful example of the improvement of a *de novo* enzyme by directed  
73 evolution has been the engineering of HG3.17,<sup>15</sup> the most active Kemp eliminase reported to  
74 date.<sup>11</sup> This artificial enzyme, which catalyzes the deprotonation of 5-nitrobenzisoxazole into the

75 corresponding *o*-cyanophenolate (Figure 1a), was evolved from the *in silico* design HG3<sup>5</sup> over 17  
76 rounds of mutagenesis and screening that also yielded the HG3.3b, HG3.7, and HG3.14  
77 intermediates (Figure 1b, Supplementary Table 1). A total of 17 mutations were introduced into  
78 HG3 during evolution to produce HG3.17, resulting in a >1000-fold increase in catalytic efficiency  
79 (Table 1, Supplementary Figure 1). Of these mutations, 11 occurred at positions within or close to  
80 the active site, including 8 at positions that were optimized during computational design of HG3  
81 (Table 1). One of the key active-site mutations occurred at position 50, which was mutated twice  
82 during evolution, first from lysine to histidine (HG3 to HG3.3b) and then from histidine to  
83 glutamine (HG3.3b to HG3.7), resulting in a novel catalytic residue ideally positioned for  
84 stabilizing negative charge buildup on the phenolic oxygen at the transition state (Figure 1a).

85 Comparison of the crystal structure of the earlier *in silico* design HG2 (Supplementary Figure 2)  
86 with that of a double mutant of HG3.17, in which surface mutations N47E and D300N were  
87 reverted to the corresponding amino acids found in HG3 to facilitate crystallization (HG3.17-  
88 E47N/N300D, PDB ID: 4BS0),<sup>15</sup> revealed that catalytic activity was also enhanced via optimized  
89 alignment of the transition-state analogue (TSA) with the catalytic base Asp127 (Figure 1c), and  
90 improved active-site complementarity to this ligand (Figure 1d). Given that subtle changes to the  
91 conformational ensemble of an enzyme can lead to significant rate enhancements,<sup>16-18</sup> it is possible  
92 that mutations in HG3.17 also contributed to enhanced catalytic efficiency by altering the  
93 conformational landscape to enrich catalytically-competent sub-states. However, the structures of  
94 HG2 and HG3.17-E47N/N300D were solved in the presence of bound TSA and at cryogenic  
95 temperatures, which could have shifted the conformational ensemble towards a single predominant  
96 sub-state, thereby limiting our ability to evaluate changes to the conformational landscape during  
97 directed evolution.

98

99 *Room-temperature crystal structures*

100 To evaluate changes to the HG3 conformational ensemble along its evolutionary trajectory,  
101 we solved room-temperature (277 K) X-ray crystal structures of all HG-series Kemp eliminases,  
102 both in the presence and absence of bound TSA. Room-temperature X-ray crystallography can  
103 reveal conformational heterogeneity in protein structures that would not be visible at cryogenic  
104 temperatures and thereby provide insights into the conformational ensemble that is sampled by a  
105 protein in solution.<sup>19</sup> All five enzymes yielded crystals under similar conditions (Supplementary  
106 Table 2), and these diffracted at resolutions of 1.35–1.99 Å (Supplementary Table 3). All unit cells  
107 corresponded to space group P<sub>2</sub>12<sub>1</sub>2<sub>1</sub> with two protein molecules in the asymmetric unit, except  
108 that of HG3.17, whose asymmetric unit was half the volume of the others and contained only one  
109 polypeptide chain although the space group was also P<sub>2</sub>12<sub>1</sub>2<sub>1</sub>. This result is in contrast with the  
110 deposited structure of HG3.17-E47N/N300D, which contains two molecules in a unit cell of  
111 identical space group and similar dimensions to those of all other HG variants reported here.<sup>15</sup> This  
112 discrepancy between our structure of HG3.17 and the previously published structure of HG3.17-  
113 E47N/N300D is likely caused by the presence of the Asn47 surface residue in all variants except  
114 for HG3.17, since this amino acid is involved in crystal packing interactions.

115 All HG-series enzymes bound the TSA in the same catalytically productive pose (Figure  
116 2a) as that observed in HG2 and HG3.17-E47N/N300D (Figure 1c–d). In this pose, the acidic N–  
117 H bond of the TSA that mimics the cleavable C–H bond of the substrate is located within  
118 hydrogen-bonding distance to the carboxylate oxygen of Asp127 (2.5–2.6 Å distance between  
119 heavy atoms), while the basic nitrogen atom corresponding to the phenolic oxygen of the transition  
120 state forms an H-bond with either a water molecule (HG3), the N<sub>e</sub> atom of His50 (HG3.3b), or the

121 side-chain amide nitrogen of Gln50 (HG3.7, HG3.14, HG3.17). In addition to being held in place  
122 by these polar interactions, the TSA is sandwiched between the hydrophobic side chains of Trp44  
123 and Met237 (Figure 2b), which are part of a binding pocket that also includes the side chains of  
124 Ala21, Met/Cys84, Met172, Leu236, Thr265, and Phe/Met267, as well as the backbone of Gly83  
125 and Pro45 (Supplementary Figure 3). Interestingly, the *cis* peptide bond formed between residues  
126 83 and 84 that is present in the original *Thermoascus aurantiacus* xylanase 10A template used to  
127 design HG3 (PDB ID: 1GOR<sup>20</sup>) is maintained in all HG structures (Figure 2c), even though both  
128 residues were mutated to obtain HG3 (H83G and T84M). In addition to adopting a *cis*  
129 conformation, which is stabilized by hydrogen bonding to an ordered water molecule, this peptide  
130 bond also adopts the *trans* conformation in the structures of TSA-bound HG3 and HG3.3b.  
131 However, starting at HG3.7, the peptide bond is found exclusively in the *cis* conformation in the  
132 TSA-bound structures because it is stabilized by an additional hydrogen bond with the Gln50 side-  
133 chain carbonyl oxygen. This hydrogen bonding interaction helps to lock Gln50 in a conformation  
134 that is properly oriented to stabilize negative charge buildup on the phenolic oxygen at the  
135 transition state, likely accounting for the majority of the 12-fold catalytic efficiency enhancement  
136 observed in HG3.7 relative to HG3.3b (Table 1).

137 From HG3.7 to HG3.17, no further changes in catalytic residues occurred during evolution.  
138 Yet, the catalytic efficiency increased by approximately 6-fold (Table 1). To evaluate whether this  
139 increase in activity was caused by changes to the conformational ensemble, we analyzed the B-  
140 factors of catalytic residues, which can be interpreted as a measure of the average displacement of  
141 an atom, or group of atoms, in the crystal. Since both conformational heterogeneity and crystalline  
142 disorder can contribute to atomic B-factors, with the latter effect potentially varying between  
143 different crystals, we calculated the Z-scores of the atomic B-factors and compared those across

144 our crystal structures of different HG variants. This Z-score analysis allowed us to evaluate the  
145 variation of B-factors relative to the mean value within an individual crystal, and showed that  
146 rigidity of the Asp127 side chain did not vary significantly during evolution (Figure 3a). By  
147 contrast, the side chain of residue 50 became increasingly rigidified over the course of the  
148 evolutionary trajectory. Increasing rigidity at position 50 is expected when this residue is mutated  
149 from a lysine to a histidine (HG3 to HG3.3b), given the lower number of degrees of freedom in  
150 the latter amino acid. This trend is also expected when histidine at position 50 is mutated to a  
151 glutamine (HG3.3b to HG3.7) given the ability of glutamine but not histidine to hydrogen-bond  
152 with the *cis* peptide formed by residues Gly83 and Cys84 (Figure 2c). However, rigidity continues  
153 to increase at this position between HG3.7 and HG3.17, even though the side-chain rotamer of  
154 Gln50 in the presence of bound TSA remains the same (Figure 2a). This result suggests that other  
155 structural features contribute to the increased rigidity observed at this position.

156 To verify the underlying cause of the increased rigidity at position 50, we calculated the  
157 average Z-score of atomic B-factors for each residue. We observed a trend whereby the loop  
158 formed by residues 87–90, which is located directly on top of residue 50, becomes increasingly  
159 rigidified during evolution (Figure 3b). Interestingly, two residues forming this loop (89 and 90)  
160 were mutated multiple times over the course of the evolutionary trajectory (Table 1). These  
161 mutations induce a conformational change in the loop that moves it closer to the active site, which  
162 results in a pi-stacking interaction between the phenyl and carboxamide groups of Phe90 and  
163 Gln50 that increases rigidity of the catalytic residue (Figure 2d, Supplementary Figure 4a).

164 A key determinant of efficient enzyme catalysis is active site pre-organization, which  
165 enables enzymes to bind substrates in a geometry close to that of the transition state. To evaluate  
166 changes in active site pre-organization during evolution, we compared the structures of HG-series

167 Kemp eliminases in the presence and absence of bound TSA. In all enzymes except for HG3.17,  
168 the unbound state is never pre-organized for catalysis as both Trp44 and Met237 adopt  
169 conformations that would prevent productive binding of the TSA (Figure 2e). In addition, the  
170 His50 and Gln50 catalytic residues in HG3.3b and HG3.7, respectively, adopt a low-occupancy,  
171 catalytically non-productive conformation in the unbound state that cannot interact favorably with  
172 the TSA. Interestingly, the non-productive conformation of Gln50 in the HG3.7 unbound state  
173 (26% occupancy) cannot stabilize the *cis* peptide bond formed by residues 83 and 84 via a  
174 hydrogen bonding interaction, and accordingly, the *trans* peptide conformation is also observed in  
175 this structure (25% occupancy) (Supplementary Figure 4b).

176 In contrast with all other HG variants, the unbound state of HG3.17 is correctly pre-  
177 organized for catalysis in a large portion of the molecules in the crystal, with only Trp44 adopting  
178 a non-productive conformation at 62% occupancy (Figure 2e). In this variant, Met237 adopts  
179 exclusively the productive conformer in the unbound state, which is stabilized by packing  
180 interactions with the neighboring Met267 side chain, a mutation that was introduced late in the  
181 evolutionary trajectory (HG3.14 to HG3.17). Overall, three of the four residues that are key for  
182 binding and stabilizing the TSA (Gln50, Asp127, Met237) adopt a catalytically productive  
183 conformation in the HG3.17 unbound state, resulting in approximately 40% of the molecules in  
184 the crystal being correctly pre-organized for efficient catalysis.

185 Enhanced complementarity to the transition state is another important feature of efficient  
186 catalysis. Therefore, computational enzyme design algorithms aim to optimize packing of the  
187 transition state. However, transition-state overpacking may reduce catalytic efficiency by creating  
188 a high-energy barrier preventing substrate entry and product release. To evaluate whether active-  
189 site accessibility changed during evolution, we calculated the active-site entrance bottleneck radius

190 on TSA-bound structures.<sup>21</sup> We observed that during evolution, the active-site bottleneck formed  
191 by the side chains of residues 50 and 267, became widened (Figure 2f), as did the mouth of the  
192 substrate entry channel formed by residues Arg275 and Trp276, which were mutated to smaller  
193 amino acids. This widening of the active site entrance could help to eliminate high-energy barriers  
194 to substrate entry and product release that could have been caused by tighter packing of the TSA  
195 in higher activity HG variants.

196

197 *HG4, an efficient artificial enzyme*

198 All of the structural features that enhance activity described above are caused primarily by  
199 residues within or close to the active site, which suggests that mutagenesis far from the active site  
200 may not be essential to create an efficient artificial enzyme. To test this hypothesis, we generated  
201 a variant of HG3 that contains all HG3.17 mutations found within 7.5 Å of the TSA, with the  
202 exception of N47E, which we omitted to favor the formation of a unit cell similar to that of HG3.  
203 We also included the R275A and W276F mutations found to widen the active site entrance. This  
204 yielded HG4, a variant of HG3 containing 8 mutations (Table 1, Supplementary Table 1). Kinetic  
205 analysis of HG4 revealed that its catalytic efficiency is >700-fold higher than that of HG3 (Table  
206 1, Supplementary Figure 1), and equivalent to that of the average natural enzyme ( $\sim 10^5 \text{ M}^{-1}\text{s}^{-1}$ ).<sup>22</sup>  
207 Crystallographic analysis of HG4 (Supplementary Tables 2–3) showed that its structure is highly  
208 similar to that of HG3.17 but with an active site that is better pre-organized (Figure 2–3,  
209 Supplementary Figure 3–4). Interestingly, both HG4 and HG3.17 have catalytic efficiencies on the  
210 order of  $10^5 \text{ M}^{-1} \text{ s}^{-1}$  despite the fact that the former enzyme contains less than half of the latter's  
211 mutations, demonstrating that distal mutations in HG3.17 contribute little to its catalytic efficiency.

212

213 Computational design of HG4

214 Given that all but one mutation (G82A) in HG4 are found at sites that were optimized  
215 during design of HG2,<sup>5</sup> we investigated whether the HG4 structure could be accurately predicted  
216 using a computational protocol similar to the one that produced HG2 (Methods, Supplementary  
217 Tables 4–6). To do so, we first performed a positive control calculation in which rotamers for the  
218 HG4 sequence were optimized on the crystal structure backbone of TSA-bound HG4. This  
219 calculation yielded an *in silico* model of HG4 with an energy score and a predicted rotameric  
220 configuration in excellent agreement with the crystal structure (Figure 4a). This control  
221 demonstrates that the combination of energy function, rotamer library, and search algorithm used  
222 in this protocol is sufficiently accurate for recapitulating the structure of HG4, provided that the  
223 correct template, binding pose, and catalytic dyad is allowed. By contrast, when we replaced the  
224 HG4 backbone template with the *Thermoascus aurantiacus* xylanase 10A backbone used to design  
225 HG2 (PDB ID: 1GOR),<sup>20</sup> we obtained a structural model that differs significantly from the HG4  
226 crystal structure and that is destabilized by approximately 40 kcal/mol (Figure 4b). This result  
227 demonstrates that the 1GOR backbone template is not well-suited to accommodate the HG4  
228 sequence, as evidenced by differences between the 1GOR-derived model and the HG4 crystal  
229 structure. Specifically, the backbone at position 83 is shifted by 1.1 Å in the HG4 crystal structure  
230 relative to its position in the 1GOR template, causing the transition state to adopt an alternate  
231 binding pose that minimizes steric clashes with Gly83, which is accompanied by repacking of  
232 several residues around the transition state, including Gln50. Use of our HG3 crystal structures  
233 with or without TSA as the design template causes similar, but less severe, structural and energetic  
234 effects (Figure 4c,d). However, when we optimized rotamers for the HG4 sequence on ensembles  
235 of backbone templates generated using molecular dynamics restrained by the HG3 diffraction data

236 (Methods), we were able to recapitulate the correct transition-state binding mode on several  
237 individual ensemble members, with energies comparable to that of the HG4 crystal structure  
238 (Figure 4e–f, Supplementary Figure 5). These results highlight the impact of small backbone  
239 geometry variations on the computational predictions, and suggest that computational enzyme  
240 design with a crystallographically-derived backbone ensemble could obviate the need for directed  
241 evolution by allowing catalytically-competent sub-states to be sampled during the design  
242 procedure.

243

## 244 **Discussion**

245 In this work, we followed changes to the conformational ensemble that occur during  
246 evolution of an enzyme with *de novo* biocatalytic function. Unlike previous examples where the  
247 active sites of *de novo* enzymes were completely remodeled during evolution,<sup>23,24</sup> or where the  
248 binding pose of the substrate or transition state analogue was significantly altered,<sup>13,17</sup> we observed  
249 only subtle changes to the active site geometry or TSA binding pose in the HG-series of Kemp  
250 eliminases. By contrast, many of the structural changes that contribute to enhanced catalysis in the  
251 HG series are dynamic in nature: the Gln50 catalytic residue became more rigid even though its  
252 average structure did not vary substantially, and the active site became better pre-organized via  
253 enrichment of catalytically-productive conformations of TSA-binding residues that were already  
254 present in the unbound state. These observations illustrate how small changes to the active site  
255 conformational ensemble can drive large changes in catalytic efficiency. Since these changes can  
256 be subtle and difficult to predict computationally, directed evolution can help increase activity by  
257 selecting for mutations that enrich catalytically-competent sub-states.<sup>17,18</sup>

258        Despite the challenges inherent to enzyme design, which are highlighted by our  
259    observations of the effects of mutations in the HG series of Kemp eliminases, our results suggest  
260    that *de novo* enzymes with native-like catalytic efficiencies can be computationally-designed,  
261    without the need to rely on subsequent improvement by laboratory directed evolution. Indeed, all  
262    mutations found in HG4 relative to the wild-type *Thermoascus aurantiacus* xylanase 10A template  
263    from which it is derived (PDB ID: 1GOR) are found at either first or second-shell residues, and  
264    these sites were all optimized during the original design of HG2.<sup>5</sup> Yet, Privett *et al.* designed the  
265    lower activity enzyme HG2 instead of HG4. While Gln50 was not sampled as part of the catalytic  
266    dyad during design of HG2, the combination of the Asp127/Gln50 dyad with the productive  
267    transition-state binding pose would have scored poorly on the 1GOR template regardless.  
268    However, our approach to computational enzyme design that utilized an experimentally-derived  
269    ensemble of backbone templates yielded HG4 models with energies and binding modes  
270    comparable to that of the HG4 crystal structure. These results suggest an iterative approach to  
271    computational enzyme design that could circumvent the need for directed evolution by introducing  
272    an additional round of design that utilizes a backbone ensemble generated from experimental  
273    structural data obtained for an initial, low-activity enzyme. In the case of evolution, mutations are  
274    not selected for in the context of a single backbone conformation but instead across an entire  
275    conformational ensemble.<sup>18</sup> Our ensemble design approach should therefore be more accurate than  
276    traditional approaches relying on a single backbone template because it allows the accessible  
277    conformational ensemble to be represented in the scoring of sequences. The incorporation of  
278    experimental restraints in the generation of the ensemble ensures that the computational procedure  
279    is applied to the true conformational ensemble that is sampled by the enzyme.

280        The results reported here provide additional support for the well-known fact that enzymes  
281        are plastic molecules whose backbone conformation can change upon introduction of mutations  
282        (as seen when comparing the 1GOR and HG-series crystal structures), and suggest improvements  
283        to the enzyme design protocol that can account for this property. This could be achieved by  
284        incorporating flexible backbone design algorithms during the repacking step,<sup>25,26</sup> or by using pre-  
285        generated ensembles of energetically-accessible backbone templates,<sup>27,28</sup> as was done here. While  
286        these methodological changes may improve the design of the enzyme transition state, it is likely  
287        that the creation of *de novo* enzymes with native-like catalytic efficiencies for more complex  
288        reactions will require a holistic approach where every possible state that the enzyme samples along  
289        its reaction coordinate is included in the design calculation. This could be achieved by the  
290        implementation of multistate approaches to computational protein design that allow the design of  
291        protein energy landscapes,<sup>29</sup> rather than single structures. We expect that the structures reported  
292        here, especially those of HG4 and HG3, will be helpful to benchmark these future enzyme design  
293        protocols.

294

## 295        **Methods**

296        *Protein expression and purification.* Codon-optimized and his-tagged (C-terminus) genes for HG-  
297        series Kemp eliminases (Supplementary Table 1) cloned into the pET-11a vector (Novagen) via  
298        *NdeI* and *BamHI* were obtained from Genscript. Enzymes were expressed in *E. coli* BL21-Gold  
299        (DE3) cells (Agilent) using lysogeny broth (LB) supplemented with 100 µg/mL ampicillin.  
300        Cultures were grown at 37 °C with shaking to an optical density at 600 nm of 0.3, at which point  
301        the incubation temperature was reduced to 18 °C. At an OD600 of 0.6, protein expression was  
302        initiated with 1 mM isopropyl β-D-1-thiogalactopyranoside. Following incubation for 16 hours at

303 18 °C with shaking (250 rpm), cells were harvested by centrifugation, resuspended in 10 mL lysis  
304 buffer (5 mM imidazole in 100 mM potassium phosphate buffer, pH 8.0), and lysed with an  
305 EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity  
306 chromatography according to the manufacturer's protocol (Qiagen), followed by gel filtration in  
307 50 mM sodium citrate buffer (pH 5.5) and 150 mM sodium chloride using an ENrich SEC 650  
308 size-exclusion chromatography column (Bio-Rad). Purified samples were concentrated using  
309 Amicon Ultracel-10K centrifugal filter units (EMD Millipore).

310

311 *Steady-state kinetics.* All assays were carried out at 27 °C in 100 mM sodium phosphate buffer  
312 (pH 7.0) supplemented with 100 mM sodium chloride. Triplicate reactions with varying  
313 concentrations of 5-nitrobenzisoxazole (AstaTech) dissolved in methanol (10% final  
314 concentration) were initiated by addition of approximately 2 µM HG3, 50 nM HG3.3b, 10 nM  
315 HG3.7/HG3.14, or 5 nM HG3.17/HG4. Product formation was monitored spectrophotometrically  
316 at 380 nm ( $\epsilon = 15,800 \text{ M}^{-1} \text{ cm}^{-1}$ ).<sup>5</sup> Linear phases of the kinetic traces were used to measure initial  
317 reaction rates. Initial reaction rates at different substrate concentrations were fit to the Michaelis-  
318 Menten equation using GraphPad Prism.

319

320 *Crystallization.* Enzyme variants were prepared in 50 mM sodium citrate buffer (pH 5.5) at the  
321 concentrations listed in Supplementary Table 2. For samples that were co-crystallized with the  
322 transition state analog (TSA) 5-nitrobenzotriazole (AstaTech), a 100 mM stock solution of the  
323 TSA was prepared in dimethylsulfoxide (DMSO) and diluted 20-fold in the enzyme solutions for  
324 a final concentration of 5 mM TSA (5% DMSO). For each enzyme variant, we carried out initial  
325 crystallization trials in 15-well hanging drop format using EasyXtal crystallization plates (Qiagen)

326 and a crystallization screen that was designed to explore the chemical space around the  
327 crystallization conditions reported by Blomberg *et al.*<sup>15</sup> Crystallization drops were prepared by  
328 mixing 1 µL of protein solution with 1 µL of the mother liquor, and sealing the drop inside a  
329 reservoir containing an additional 500 µL of the mother liquor solution. The mother liquor  
330 solutions contained ammonium sulfate as a precipitant in sodium acetate buffer (100 mM), and the  
331 specific growth conditions that yielded the crystals used for X-ray data collection are provided in  
332 Supplementary Table 2. In some cases, a microseeding protocol was required to obtain high-  
333 quality crystals. Microseeds were prepared by vortexing crystals in their mother liquor in the  
334 presence of glass beads (0.5 mm), and were subsequently diluted into the mother liquor solutions  
335 used to form the crystallization drops.

336

337 *X-ray data collection and processing.* Prior to X-ray data collection, crystals were mounted in  
338 polyimide loops and sealed using a MicroRT tubing kit (MiTeGen). Single-crystal X-ray  
339 diffraction data was collected on beamline 8.3.1 at the Advanced Light Source. The beamline was  
340 equipped with a Pilatus3 S 6M detector, and was operated at a photon energy of 11111 eV. Crystals  
341 were maintained at 277 K throughout the course of data collection. Each data set was collected  
342 using a total X-ray dose of 200 kGy or less, and covered a 180° wedge of reciprocal space. Multiple  
343 data sets were collected for each enzyme variant.

344 X-ray data was processed with the Xia2 program  
345 (<https://doi.org/10.1107/S0021889809045701>), which performed indexing, integration, and  
346 scaling with XDS and XSCALE<sup>30</sup> followed by merging with Pointless.<sup>31</sup> For each variant,  
347 multiple individual data sets were merged to obtain the final set of reduced intensities, and the  
348 resolution cutoff was taken where the CC<sub>1/2</sub> and <I/σI> values for the merged intensities fell to

349 approximately 0.5 and 1.0 respectively. Information regarding data collection and processing is  
350 presented in Supplementary Table 3. The reduced diffraction data were analyzed with  
351 phenix.xtriage ([http://www ccp4.ac.uk/newsletters/newsletter43/articles/PHZ\\_RWGK\\_PDA.pdf](http://www ccp4.ac.uk/newsletters/newsletter43/articles/PHZ_RWGK_PDA.pdf))  
352 to check for crystal pathologies, and no complications were identified.

353

354 *Structure determination.* We obtained initial phase information for calculation of electron density  
355 maps by molecular replacement using the program Phaser,<sup>32</sup> as implemented in the PHENIX  
356 suite.<sup>33</sup> Several different HG-series enzymes were used as molecular replacement search models.  
357 All members of the HG-series of enzymes crystallized in the same crystal form, containing two  
358 copies of the molecule in the crystallographic asymmetric unit, except for HG3.17, which  
359 crystallized with only one molecule in the asymmetric unit. To avoid model bias that could  
360 originate from using other members of the HG-series as molecular replacement search models, we  
361 applied random coordinate displacements ( $\sigma = 0.5 \text{ \AA}$ ) to the atoms, and performed coordinate  
362 refinement against the structure factor data before proceeding to manual model building.

363 Next, we performed iterative steps of manual model rebuilding followed by refinement of  
364 atomic positions, atomic displacement parameters, and occupancies using a translation-libration-  
365 screw (TLS) model, a riding hydrogen model, and automatic weight optimization. All model  
366 building was performed using Coot<sup>34</sup> and refinement steps were performed with phenix.refine  
367 (v1.13-2998) within the PHENIX suite.<sup>33,35</sup> Restraints for the TSA were generated using  
368 phenix.elbow,<sup>36</sup> starting from coordinates available in the Protein Data Bank (PDB ligand ID:  
369 6NT).<sup>37</sup> Further information regarding model building and refinement, as well as PDB accession  
370 codes for the final models, are presented in Supplementary Table 3. Time-averaged ensembles  
371 were generated for HG3 with and without ligand with phenix.ensemble\_refinement implemented

372 in PHENIX. To prepare the structures for ensemble refinement, low-occupancy conformers were  
373 removed, and occupancies adjusted to 100% using phenix.pdbtools. Hydrogen atoms were then  
374 added using phenix.ready\_set. This procedure yielded 79- and 49-member ensembles from the  
375 HG3 structures with and without TSA, respectively.

376

377 *Computational enzyme design.* All calculations were performed with the Triad protein design  
378 software (Protabit, Pasadena, CA, USA) using a Monte Carlo with simulated annealing search  
379 algorithm for rotamer optimization. The crystal structure of *Thermoascus aurantiacus* xylanase  
380 10A was obtained from the Protein Data Bank (PDB code: 1GOR<sup>20</sup>) and further refined as  
381 described above to fix modeling issues with Thr84. Structures of HG3 with and without TSA, HG4  
382 with TSA, and ensembles of HG3-derived templates were obtained from refinement of  
383 crystallographic data as described above. Following extraction of protein heavy-atom coordinates  
384 for the highest occupancy conformer from chain A, hydrogen atoms were added using the *addH.py*  
385 application in Triad. The Kemp elimination transition state (TS) structure<sup>38</sup> was built using the  
386 parameters described by Privett and coworkers.<sup>5</sup> Residue positions surrounding Asp127 were  
387 mutated to Gly (Supplementary Table 4), with the exception of position 50, which was mutated to  
388 Gln. A backbone-independent rotamer library<sup>39</sup> with expansions of  $\pm 1$  standard deviation around  
389  $\chi_1$  and  $\chi_2$  was used to provide side-chain conformations. A library of TS poses was generated in  
390 the active site by targeted ligand placement<sup>2</sup> using the contact geometries listed in Supplementary  
391 Table 5. TS pose energies were calculated using the PHOENIX energy function,<sup>5</sup> which consists  
392 of a Lennard-Jones 12–6 van der Waals term from the Dreiding II force field<sup>40</sup> with atomic radii  
393 scaled by 0.9, a direction-dependent hydrogen bond term with a well depth of 8.0 kcal mol<sup>-1</sup> and  
394 an equilibrium donor-acceptor distance of 2.8 Å,<sup>41</sup> an electrostatic energy term modelled using

395 Coulomb's law with a distance-dependent dielectric of 10, an occlusion-based solvation potential  
396 with scale factors of 0.05 for nonpolar burial, 2.5 for nonpolar exposure, and 1.0 for polar burial,<sup>42</sup>  
397 and a secondary structural propensity term.<sup>43</sup> During the energy calculation step, TS–side-chain  
398 interaction energies were biased to favor interactions that satisfy contact geometries  
399 (Supplementary Table 6) as described by Lassila *et al.*<sup>2</sup>

400 Following ligand placement, the 10 lowest energy TS poses found on each template (HG4  
401 with TSA, 1GOR, HG3 with TSA, and HG3 without TSA) were selected as starting points for  
402 repacking of the HG4 sequence. For individual members of the crystallographically-derived  
403 ensembles, only the single lowest energy TS pose was used for repacking. In the repacking  
404 calculation, the TS structure was translated  $\pm 0.4$  Å in each Cartesian coordinate in 0.2-Å steps,  
405 and rotated 10° about all three axes (origin at TS geometric center) in 5° steps for a total  
406 combinatorial rotation/translation search size of 5<sup>6</sup> or 15,625 poses. Residues that were converted  
407 to Gly in the ligand placement step were allowed to sample all conformations of the amino acid  
408 found at that position in the HG4 sequence (Supplementary Table 4). The identities of the catalytic  
409 residues were fixed and allowed to sample all conformations of that amino-acid type. Side-chain–  
410 TS interaction energies were biased to favor those contacts that satisfy the geometries as done  
411 during the ligand placement step (Supplementary Table 6). Rotamer optimization was carried out  
412 using the search algorithm, rotamer library, and energy function described above. The single  
413 lowest energy repacked structure on each backbone template was used for analysis. To compare  
414 energies of the HG4 models obtained on the various templates, we calculated the energy difference  
415 between each repacked structure and the corresponding all-Gly structure obtained after ligand  
416 placement, and these energies are reported throughout the figures and text.

417

418 *Statistics and reproducibility.* Experiments were repeated in triplicate where feasible. All  
419 replications were successful and the resulting data is presented with error values representing the  
420 standard deviation between replicates. No data was excluded from analyses.

421

422 **Data availability**

423 Structure coordinates for all HG-series Kemp eliminases have been deposited in the Protein Data  
424 Bank with the following accession codes: HG3 (PDB ID: 5RG4, 5RGA), HG3.3b (PDB ID: 5RG5,  
425 5RGB), HG3.7 (PDB ID: 5RG6, 5RGC), HG3.14 (PDB ID: 5RG7, 5RGD), HG3.17 (PDB ID:  
426 5RG8, 5RGE), and HG4 (PDB ID: 5RG9, 5RGF).

427

428 **Acknowledgments**

429 R.A.C. acknowledges an Early Researcher Award from the Ontario Ministry of Economic  
430 Development & Innovation (ER14-10-139), and grants from the Natural Sciences and Engineering  
431 Research Council of Canada (NSERC, RGPIN-2016-04831) and the Canada Foundation for  
432 Innovation (26503). J.S.F. is supported by a Searle Scholar Award from the Kinship Foundation,  
433 a Pew Scholar Award from the Pew Charitable Trusts, a Packard Fellowship from the David and  
434 Lucile Packard Foundation, NIH GM110580, UC Office of the President Laboratory Fees  
435 Research Program LFR-17-476732, and NSF STC-1231306. A.B. is the recipient of a postdoctoral  
436 fellowship from NSERC. M.C.T. is supported by a Ruth L. Kirschstein National Research Service  
437 Award (F32 HL129989). We thank Dr. Heidi K. Privett for providing HG2 design scripts and  
438 computational model, and Dr. Justin T. Biel for assistance with initial processing of X-ray datasets.

439

440 Data collection at Beamline 8.3.1 at the Advanced Light Sources is supported by the University of  
441 California Office of the President, Multicampus Research Programs and Initiatives grant MR-15-  
442 328599, the Program Breakthrough Biomedical Research (which is partially funded by the Sandler  
443 Foundation), the National Institutes of Health (R01 GM124149 and P30 GM124169), Plexxikon  
444 Inc., and the Integrated Diffraction Analysis Technologies program of the US Department of  
445 Energy Office of Biological and Environmental Research. The Advanced Light Source (Berkeley,  
446 CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of  
447 the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic  
448 Energy Sciences.

449

450 **Author Contributions**

451 A.B. and R.A.C. conceived the project. A.B., N.Z., E.N., N.M., and L.L. purified proteins. A.B.  
452 and N.Z. performed enzyme kinetics experiments. R.A.C. and M.C.T. crystallized proteins and  
453 performed X-ray diffraction experiments. A.B., R.V.R., M.C.T., and R.A.C. performed  
454 refinements. M.C.T. and J.S.F. designed X-ray crystallography experiments. A.B. and R.V.R.  
455 performed computational design experiments. R.A.C. wrote the manuscript. A.B., R.V.R., and  
456 M.C.T. edited the manuscript.

457

458 **Competing Interests**

459 The authors declare no competing interests.

460 **References**

- 461 (1) Edwards, D. R.; Lohman, D. C.; Wolfenden, R. Catalytic proficiency: the extreme case of  
462 S-O cleaving sulfatases *J Am Chem Soc* **134**,525-531 (2012).
- 463 (2) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. Combinatorial methods for small-  
464 molecule placement in computational enzyme design *Proc Natl Acad Sci U S A* **103**,16710-16715 (2006).
- 465 (3) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.;  
466 Rothlisberger, D.; Baker, D. New algorithms and an in silico benchmark for computational enzyme design  
467 *Protein Sci* **15**,2785-2794 (2006).
- 468 (4) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.;  
469 Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp  
470 elimination catalysts by computational enzyme design *Nature* **453**,190-195 (2008).
- 471 (5) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.;  
472 Houk, K. N.; Mayo, S. L. Iterative approach to computational enzyme design *Proc Natl Acad Sci U S A*  
473 **109**,3790-3795 (2012).
- 474 (6) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.;  
475 Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., 3rd; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D.  
476 De novo computational design of retro-aldol enzymes *Science* **319**,1387-1391 (2008).
- 477 (7) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z.; Smith, M.; Hari, S.;  
478 Kast, P.; Herschlag, D.; Hilvert, D.; Baker, D. Robust design and optimization of retroaldol enzymes  
479 *Protein Sci* **21**,717-726 (2012).
- 480 (8) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher,  
481 J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational design  
482 of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction *Science* **329**,309-313 (2010).
- 483 (9) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J.; Gallaher, J.;  
484 Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.;  
485 Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D.  
486 Computational design of catalytic dyads and oxyanion holes for ester hydrolysis *J Am Chem Soc*  
487 **134**,16197-16206 (2012).
- 488 (10) Bjelic, S.; Nivón, L. G.; Çelebi-Ölçüm, N.; Kiss, G.; Rosewall, C. F.; Lovick, H. M.; Ingalls, E.  
489 L.; Gallaher, J. L.; Seetharaman, J.; Lew, S.; Montelione, G. T.; Hunt, J. F.; Michael, F. E.; Houk, K. N.;  
490 Baker, D. Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-  
491 Hillman reaction *ACS Chem Biol* **8**,749-757 (2013).
- 492 (11) St-Jacques, A. D.; Gagnon, O.; Chica, R. A. In *Modern Biocatalysis: Advances Towards  
493 Synthetic Biological Systems*; The Royal Society of Chemistry: 2018, p 88-116.
- 494 (12) Mak, W. S.; Siegel, J. B. Computational enzyme design: transitioning from catalytic  
495 proteins to enzymes *Curr Opin Struct Biol* **27**,87-94 (2014).
- 496 (13) Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.;  
497 Tawfik, D. S. Bridging the gaps in design methodologies by evolutionary optimization of the stability and  
498 proficiency of designed Kemp eliminase KE59 *Proc Natl Acad Sci U S A* **109**,10358-10363 (2012).
- 499 (14) Khersonsky, O.; Rothlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss,  
500 G.; Houk, K. N.; Baker, D.; Tawfik, D. S. Optimization of the in-silico-designed kemp eliminase KE70 by  
501 computational design and directed evolution *J Mol Biol* **407**,391-412 (2011).
- 502 (15) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R.; Grutter, M. G.; Privett, H. K.; Mayo, S.  
503 L.; Hilvert, D. Precision is essential for efficient catalysis in an evolved Kemp eliminase *Nature* **503**,418-  
504 421 (2013).
- 505 (16) Romero-Rivera, A.; Garcia-Borras, M.; Osuna, S. Role of Conformational Dynamics in the  
506 Evolution of Retro-Aldolase Activity *Acs Catalysis* **7**,8524-8532 (2017).

- 507 (17) Hong, N. S.; Petrovic, D.; Lee, R.; Gryn'ova, G.; Purg, M.; Saunders, J.; Bauer, P.; Carr, P.  
508 D.; Lin, C. Y.; Mabbitt, P. D.; Zhang, W.; Altamore, T.; Easton, C.; Coote, M. L.; Kamerlin, S. C. L.; Jackson,  
509 C. J. The evolution of multiple active site configurations in a designed enzyme *Nat Commun* **9**,3900  
510 (**2018**).  
511 (18) Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.;  
512 Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J. The role of protein  
513 dynamics in the evolution of new enzyme function *Nat Chem Biol* **12**,944-950 (**2016**).  
514 (19) Biel, J. T.; Thompson, M. C.; Cunningham, C. N.; Corn, J. E.; Fraser, J. S. Flexibility and  
515 Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin  
516 *Structure* **25**,739-749 e733 (**2017**).  
517 (20) Lo Leggio, L.; Kalogiannis, S.; Eckert, K.; Teixeira, S. C.; Bhat, M. K.; Andrei, C.; Pickersgill,  
518 R. W.; Larsen, S. Substrate specificity and subsite mobility in *T. aurantiacus* xylanase 10A *FEBS Lett*  
519 **509**,303-308 (**2001**).  
520 (21) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.;  
521 Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: a tool for the  
522 analysis of transport pathways in dynamic protein structures *PLoS Comput Biol* **8**,e1002708 (**2012**).  
523 (22) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R. The  
524 moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters  
525 *Biochemistry* **50**,4402-4410 (**2011**).  
526 (23) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a  
527 designed retro-alcoholase leads to complete active site remodeling *Nat Chem Biol* **9**,494-498 (**2013**).  
528 (24) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R.; Baker, D.; Griffiths, A. D.; Hilvert, D.  
529 Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase *Nat Chem* **9**,50-56  
530 (**2017**).  
531 (25) Friedland, G. D.; Linares, A. J.; Smith, C. A.; Kortemme, T. A simple model of backbone  
532 flexibility improves modeling of side-chain conformational variability *J Mol Biol* **380**,757-774 (**2008**).  
533 (26) Babor, M.; Mandell, D. J.; Kortemme, T. Assessment of flexible backbone protein design  
534 methods for sequence library prediction in the therapeutic antibody Herceptin-HER2 interface *Protein*  
535 *Sci* **20**,1082-1089 (**2011**).  
536 (27) Davey, J. A.; Chica, R. A. Improving the accuracy of protein stability predictions with  
537 multistate design using a variety of backbone ensembles *Proteins* **82**,771-784 (**2014**).  
538 (28) Davey, J. A.; Chica, R. A. Multistate Computational Protein Design with Backbone  
539 Ensembles *Methods Mol Biol* **1529**,161-179 (**2017**).  
540 (29) Davey, J. A.; Damry, A. M.; Goto, N. K.; Chica, R. A. Rational design of proteins that  
541 exchange on functional timescales *Nat Chem Biol* **13**,1280-1285 (**2017**).  
542 (30) Kabsch, W. Xds *Acta Crystallogr D Biol Crystallogr* **66**,125-132 (**2010**).  
543 (31) Evans, P. Scaling and assessment of data quality *Acta Crystallogr D Biol Crystallogr*  
544 **62**,72-82 (**2006**).  
545 (32) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R.  
546 J. Phaser crystallographic software *J Appl Crystallogr* **40**,658-674 (**2007**).  
547 (33) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J.  
548 J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R.  
549 J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-  
550 based system for macromolecular structure solution *Acta Crystallogr D Biol Crystallogr* **66**,213-221  
551 (**2010**).  
552 (34) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot  
553 *Acta Crystallogr D Biol Crystallogr* **66**,486-501 (**2010**).

- 554 (35) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.;  
555 Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards automated  
556 crystallographic structure refinement with phenix.refine *Acta Crystallogr D Biol Crystallogr* **68**,352-367  
557 (2012).
- 558 (36) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. electronic Ligand Builder and  
559 Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation *Acta Crystallogr*  
560 *D Biol Crystallogr* **65**,1074-1080 (2009).
- 561 (37) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.;  
562 Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for  
563 macromolecular structures *J Mol Biol* **112**,535-542 (1977).
- 564 (38) Na, J.; Houk, K. N.; Hilvert, D. Transition state of the base-promoted ring-opening of  
565 isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody  
566 production *Journal of the American Chemical Society* **118**,6462-6471 (1996).
- 567 (39) Dunbrack, R. L.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer  
568 preferences *Protein Science* **6**,1661-1681 (1997).
- 569 (40) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. Dreiding - a Generic Force-Field for Molecular  
570 Simulations *Journal of Physical Chemistry* **94**,8897-8909 (1990).
- 571 (41) Dahiyat, B. I.; Mayo, S. L. Probing the role of packing specificity in protein design *Proc  
572 Natl Acad Sci U S A* **94**,10172-10177 (1997).
- 573 (42) Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L. Generation of longer emission  
574 wavelength red fluorescent proteins using computationally designed libraries *Proc Natl Acad Sci U S A*  
575 **107**,20257-20262 (2010).
- 576 (43) Shortle, D. Propensities, Probabilities, and the Boltzmann Hypothesis *Protein Sci*  
577 **12**,1298-1302 (2003).
- 578 (44) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. Close agreement between the  
579 orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical  
580 calculations *Proc Natl Acad Sci U S A* **101**,6946-6951 (2004).

581

582 **Table 1. Kinetic parameters of Kemp eliminases**

583

Enzyme	Mutations from HG3 <sup>a</sup>	$k_{\text{cat}} / K_M$ (M <sup>-1</sup> s <sup>-1</sup> ) <sup>b</sup>
HG3	-	160 ± 10 (1300)
HG3.3b	V6I <i>K50H M84C S89R Q90D A125N</i>	2300 ± 800 (5400)
HG3.7	V6I Q37K <i>K50Q M84C S89R Q90H A125N</i>	28000 ± 4000 (37000)
HG3.14	V6I Q37K <i>K50Q G82A M84C Q90H T105I A125T T142N T208M T279S D300N</i>	55000 ± 4000 (70000)
HG3.17	V6I Q37K <i>N47E K50Q G82A M84C S89N Q90F T105I A125T T142N T208M F267M W275A R276F T279S D300N</i>	170000 ± 20000 (230000)
HG4	<i>K50Q G82A M84C Q90F A125T F267M W275A R276F</i>	120000 ± 10000

584

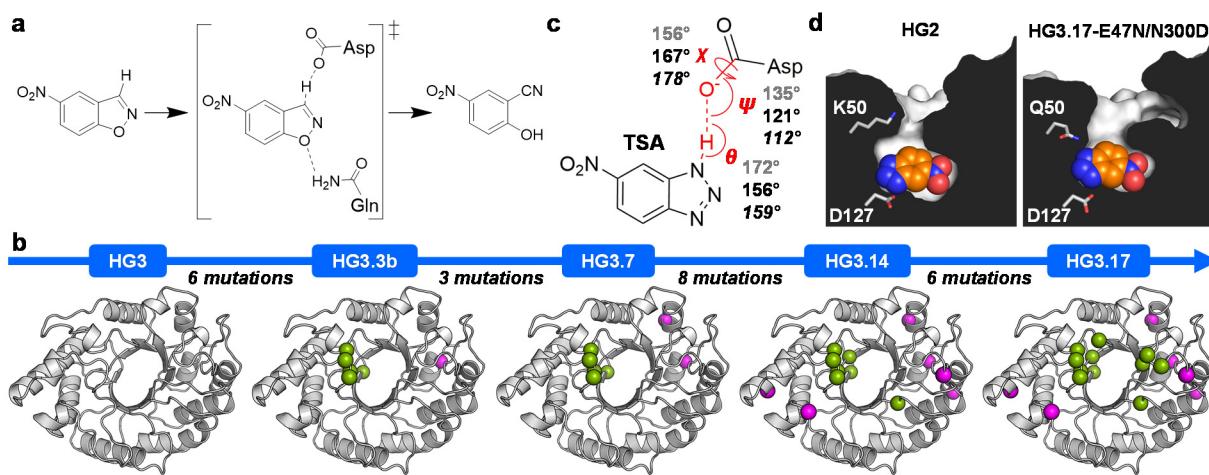
585 <sup>a</sup>Mutations in italics occurred at sites optimized during computational design of HG3.<sup>5</sup>

586 <sup>b</sup>Individual parameters  $K_M$  and  $k_{\text{cat}}$  could not be determined accurately because saturation was not possible at the  
587 maximum substrate concentration tested (2 mM). Experiments were performed in triplicate using enzymes from at  
588 least two independent enzyme preparations (mean ± s.d.). Values in parentheses are from Blomberg *et al.*<sup>15</sup>

589

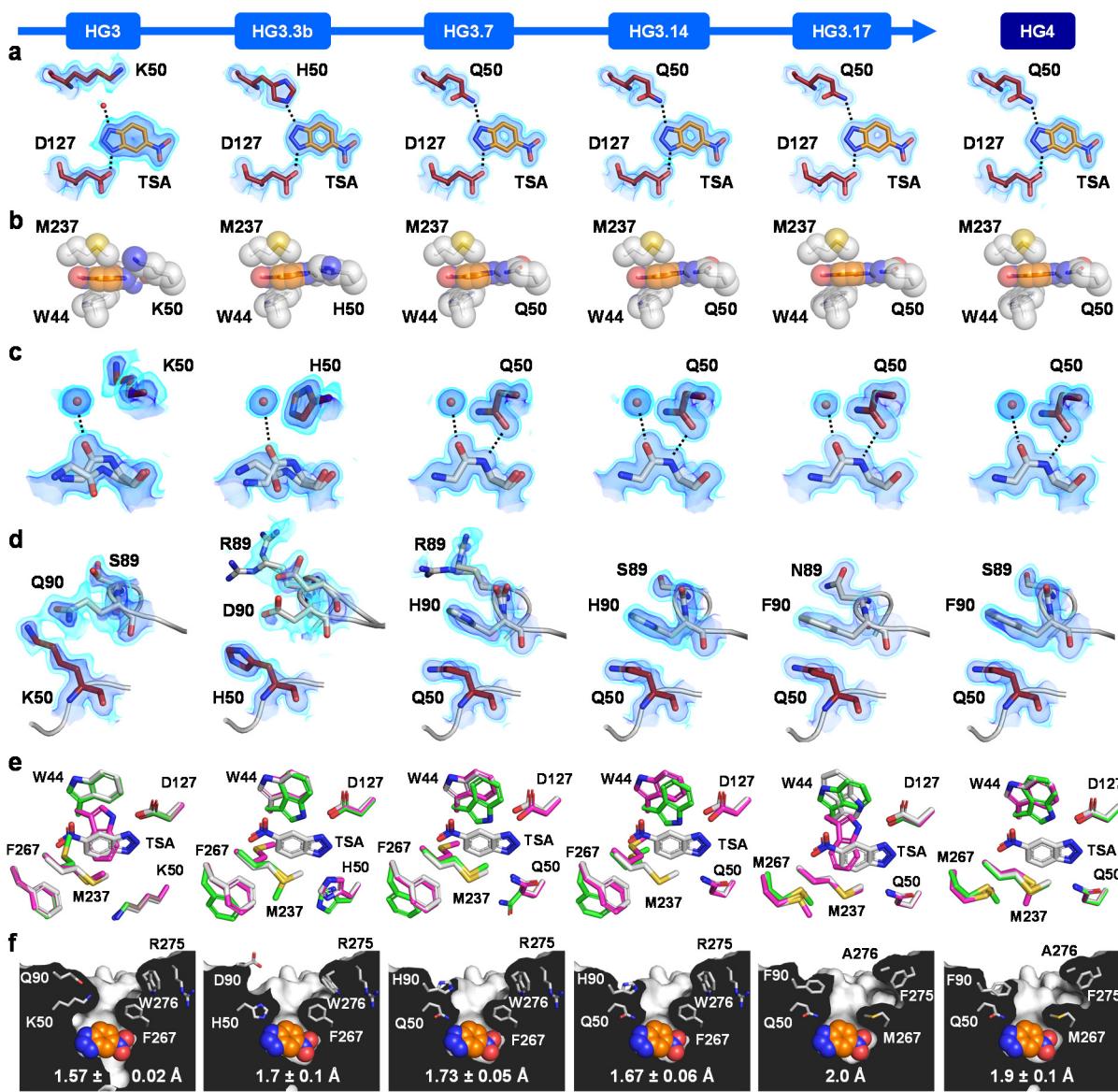
590 **Figures**

591



593 **Figure 1. HG series of Kemp eliminases.** (a) HG enzymes catalyze the Kemp elimination reaction using a catalytic  
594 dyad consisting of a base (Asp127) that deprotonates 5-nitrobenzisoxazole, and an H-bond donor (Gln50) that  
595 stabilizes negative charge buildup on the phenolic oxygen at the transition state ( $\ddagger$ ). This reaction yields 4-nitro-2-  
596 cyanophenol. (b) Directed evolution of the *in silico* design HG3, which itself is a single mutant (S265T) of the earlier  
597 design HG2. A total of 17 mutations (shown as spheres) were introduced during evolution, including 11 at positions  
598 within or close to the active site (green) and 6 at distal sites (magenta). (c) Angles describing the hydrogen bonding  
599 interaction between the transition state analog (TSA) and Asp127 in the HG2 (PDB ID: 3NYD)<sup>5</sup> and HG3.17-  
600 E47N/N300D (PDB ID: 4BSO)<sup>15</sup> crystal structures are indicated in grey and black, respectively. Values in italics are  
601 optimal angles calculated for hydrogen bonding interactions between acetamide dimers.<sup>44</sup> (d) Cut-away view of the  
602 active site pocket shows that its structural complementarity with the TSA (spheres) is improved in the higher activity  
603 variant HG3.17-E47N/N300D. Key active site residues are shown as sticks.

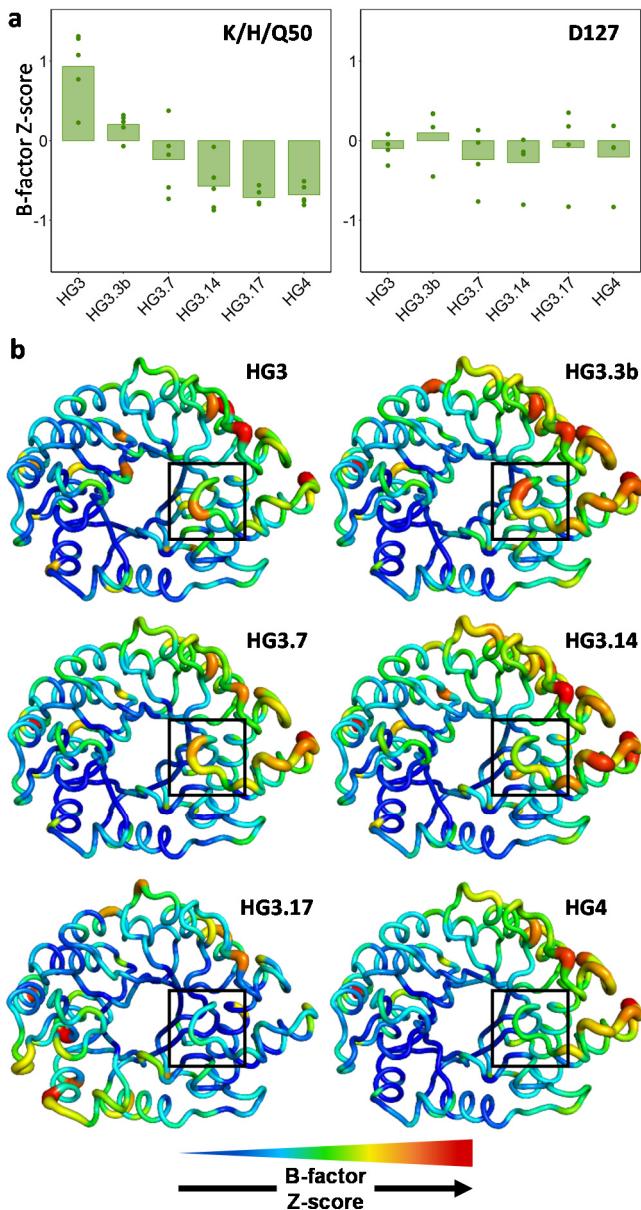
604



605  
606

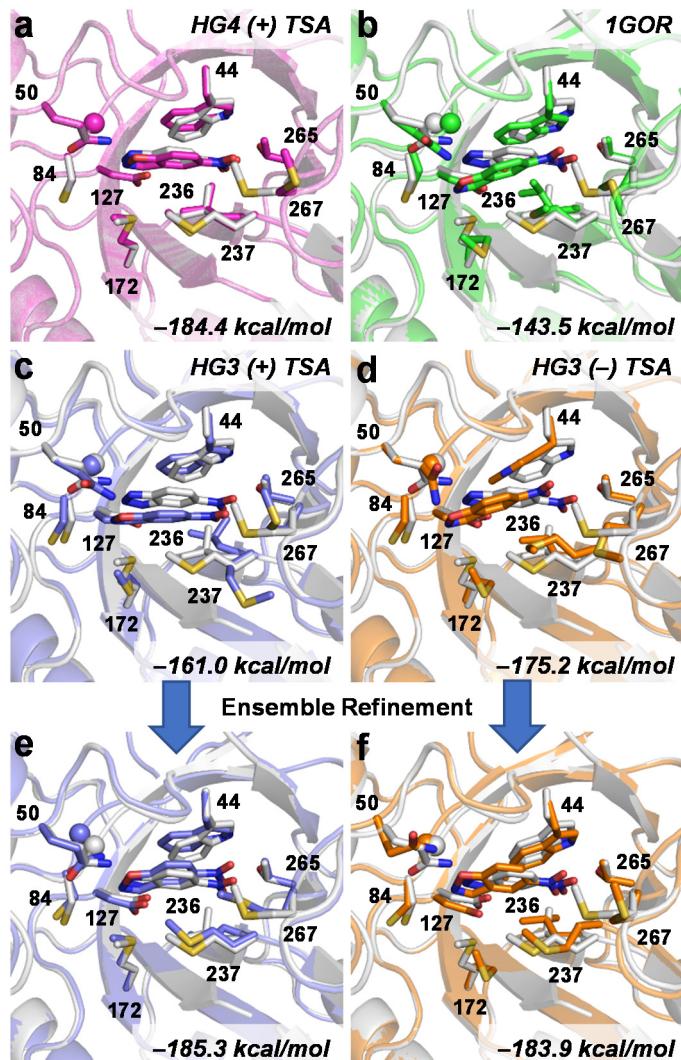
**Figure 2. Crystal structures of HG-series Kemp eliminases.** In all cases, only atoms from chain A are shown. (a) Binding pose of the transition state analogue (TSA, orange). Hydrogen bonds are shown as dashed lines. The red sphere represents a water molecule. The 2Fo-Fc map is shown in volume representation at two contour levels:  $0.5 \text{ eÅ}^{-3}$  and  $1.5 \text{ eÅ}^{-3}$  in light and dark blue, respectively. (b) The TSA (orange) is sandwiched between the hydrophobic side chains of Trp44 and Met237. (c) The peptide bond between residues 83 and 84 can adopt *cis* or *trans* conformations. Hydrogen bonds are shown as dashed lines. The 2Fo-Fc map is shown in volume representation at two contour levels:  $0.5 \text{ eÅ}^{-3}$  and  $1.5 \text{ eÅ}^{-3}$  in light and dark blue, respectively. (d) Conformational changes to loop formed by residues 87–90 over the course of the evolutionary trajectory. The 2Fo-Fc map is shown in volume representation at two contour levels:  $0.50 \text{ eÅ}^{-3}$  and  $1.5 \text{ eÅ}^{-3}$  in light and dark blue, respectively. (e) Superposition of the TSA-bound structure (white) with the highest (magenta) and lowest (green) occupancy conformers of the unbound structure for each Kemp eliminase. The occupancies of non-productive conformers of Trp44 in the unbound structures of HG3.17 and HG4 are 62% and 26%, respectively. (f) Cut-away view of the active site shows that its entrance (top) becomes widened during evolution, as indicated by an increasing bottleneck radius (reported as the average radius  $\pm$  s.d. calculated using the highest occupancy conformers from both chain A and B, except for HG3.17, which contains a single chain). The TSA is shown as orange spheres. Bottleneck radii were calculated using the PyMOL plugin Caver 3.0.<sup>21</sup>

622



623  
624

625 **Figure 3. Conformational heterogeneity.** (a) B-factor Z-scores for the residue at position 50 in the absence of bound  
626 TSA decrease over the course of the evolutionary trajectory, while those for Asp127 do not change significantly. Z-  
627 scores of individual side-chain atoms are shown as dots, while the average is indicated by the bar. Positive and negative  
628 Z-scores indicate increased flexibility or rigidity relative to the average residue in the protein, respectively. (b) B-  
629 factor Z-scores for all protein residues in the absence of bound TSA plotted on a model backbone for each Kemp  
630 eliminase. Thickness of the sausage plot increases with the B-factor Z-score, indicating increased flexibility. The loop  
631 formed by residues 87–90 (boxed) becomes more rigid during evolution.  
632



633  
634  
635 **Figure 4. Computational design of HG4 on various backbone templates.** The HG4 crystal structure with bound  
636 TSA (white) is overlaid on the HG4 design models (colored) obtained using the crystal structure of (a) HG4 with  
637 bound TSA, (b) *Thermoascus aurantiacus* xylanase 10A (PDB ID: 1GOR), (c) HG3 with bound TSA, or (d) HG3  
638 without TSA. Panels (e) and (f) show the HG4 design models obtained using the template prepared by ensemble  
639 refinement from the corresponding HG3 density map that gave the best energy following repacking. PHOENIX  
640 energies of design models after repacking are indicated at bottom right. For reference, the energy of the HG4 crystal  
641 structure with bound transition state is -182.7 kcal/mol. In all cases, the transition state and transition state analogue  
642 are show at the center of barrel. Side chains of all residues forming the binding pocket are shown with the exception  
643 of Ala21 and Pro45, which were omitted for clarity. The sphere shows the alpha carbon of Gly83.  
644

Supplementary Materials for

**Evolution of an enzyme conformational ensemble guides design of an efficient biocatalyst**

**Aron Broom,<sup>1,†</sup> Rojo V. Rakotoharisoa,<sup>1,†</sup> Michael C. Thompson,<sup>2</sup> Niayesh Zarifi,<sup>1</sup> Erin Nguyen,<sup>1</sup> Nurzhan Mukhametzhannov,<sup>1</sup> Lin Liu,<sup>2</sup> James S. Fraser<sup>2</sup> & Roberto A. Chica<sup>1,\*</sup>**

<sup>1</sup> Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie, K1N 6N5, Ottawa, Ontario, Canada

<sup>2</sup> Department of Bioengineering and Therapeutic Science, University of California, San Francisco, San Francisco, California 94158, United States

<sup>†</sup> These authors contributed equally to this work.

\* Correspondence to: E-mail: [rchica@uottawa.ca](mailto:rchica@uottawa.ca)

**This file includes:**

Supplementary Tables 1–6  
Supplementary Figures 1–5

**Supplementary Table 1.** Amino-acid sequences of HG-series Kemp eliminases

Enzyme	# mutations from HG3	Sequence <sup>a</sup>
<b>HG3</b>	–	MAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> QADFGMVWPENS MKWDATEPSQGNFNFAGADYLVNWAQQNGKLIGGGMLVWHSQPLPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>W</b> DVVGEAFNEDGSLRQTVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQTHLSAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETVFGVADPDSWRASSTPPLLFDGNFNPKPAYNAIV QDLQQGSIEGRGHHHHHH
<b>HG3.3b</b>	6	MAEAAQSI <b>D</b> QLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> QADFGMVWPENS <b>M</b> HWDATEPSQGNFNFAGADYLVNWAQQNGKLIGGGCLVWHR <b>D</b> LPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>W</b> DVVGEAFNEDGSLRQTVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQTHLSAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETVFGVADPDSWRASSTPPLLFDGNFNPKPAYNAIV QDLQQGSIEGRGHHHHHH
<b>HG3.7</b>	7	MAEAAQSI <b>D</b> QLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> KADFGMVWPENS <b>M</b> QWDATEPSQGNFNFAGADYLVNWAQQNGKLIGGGCLVWHR <b>H</b> LPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>W</b> DVVGEAFNEDGSLRQTVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQTHLSAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETVFGVADPDSWRASSTPPLLFDGNFNPKPAYNAIV QDLQQGSIEGRGHHHHHH
<b>HG3.14</b>	12	MAEAAQSI <b>D</b> QLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> KADFGMVWPENS <b>M</b> QWDATEPSQGNFNFAGADYLVNWAQQNGKLIG <b>A</b> GCLVWH <b>S</b> HLPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>T</b> DVVGEAFNEDGSLRQ <b>N</b> TVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQ <b>M</b> HLISAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETVFGVADPDS <b>A</b> FASTPPLLFDGNFNPKPAYNAIV <b>Q</b> NLQQGSIEGRGHHHHHH
<b>HG3.17</b>	17	MAEAAQSI <b>D</b> QLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> KADFGMVWP <b>E</b> S <b>M</b> QWDATEPSQGNFNFAGADYLVNWAQQNGKLIG <b>A</b> GCLVWH <b>N</b> FHLPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>T</b> DVVGEAFNEDGSLRQ <b>N</b> TVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQ <b>M</b> HLISAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETV <b>M</b> GVADPDS <b>A</b> FASTPPLLFDGNFNPKPAYNAIV <b>Q</b> NLQQGSIEGRGHHHHHH
<b>HG4</b>	8	MAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAI <b>I</b> QADFGMVWPENS MQWDATEPSQGNFNFAGADYLVNWAQQNGKLIG <b>A</b> GCLVWH <b>S</b> FLPSWVSSI TDKNLTLNVMKNHITTLMTTRYKGKIR <b>T</b> DVVGEAFNEDGSLRQTVFLNVI GEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI <b>V</b> NRVKQWRAAG VPIDGIGSQTHLSAGQQAGVLQALPLIASAGTPEVSILMLDVAGASPTDY VNVVNACLNVQSCVGETV <b>M</b> GVADPDS <b>A</b> FASTPPLLFDGNFNPKPAYNAIV QDLQQGSIEGRGHHHHHH

<sup>a</sup> Mutations from HG3 are highlighted in bold. All sequences contain a His-tag at the C-terminus.

**Supplementary Table 2.** Crystallization conditions

Enzyme <sup>a</sup>	TSA <sup>b</sup>	pH	Protein (mg mL <sup>-1</sup> )	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (M)
<b>HG3</b>	(-)	4.6	4	1.8
	(+)	4.6	4	1.8
<b>HG3.3b</b>	(-)	5.4	5	2.0
	(+)	5.4	5	2.0
<b>HG3.7</b>	(-)	4.0	6	1.6
	(+)	5.0	12	2.0
<b>HG3.14</b>	(-)	4.5	3	1.6
	(+)	5.0	3	2.4
<b>HG3.17</b>	(-)	5.0	5	1.6
	(+)	5.0	10	1.6
<b>HG4</b>	(-)	4.6	10	1.6
	(+)	4.6	10	1.6

<sup>a</sup> All enzymes were crystallized in 100 mM sodium acetate buffer at the indicated pH.

<sup>b</sup> TSA (5-nitrobenzotriazole) was dissolved in pure DMSO and added to crystallization solution at a final concentration of 5 mM (5% DMSO).

**Supplementary Table 3.** Crystallographic data and refinement statistics for room-temperature structures (277 K)

TSA	HG3		HG3.3b		HG3.7		HG3.14		HG3.17		HG4	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
PDB ID	5RG4	5RGA	5RG5	5RGB	5RG6	5RGC	5RG7	5RGD	5RG8	5RGE	5RG9	5RGF
<b>Data collection<sup>a</sup></b>												
Resolution (Å)	41.52– 1.99	55.15– 1.89	37.09– 1.62	48.16– 1.44	60.34– 1.35	79.81– 1.39	79.83– 1.47	48.94– 1.40	46.17– 1.73	38.46– 1.77	79.99– 1.47	41.50– 1.40
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>											
<i>Cell params.</i>												
a b c (Å)	76.14 79.97 99.06	76.24 79.88 99.05	76.23 80.03 98.91	76.24 79.92 98.71	76.20 79.85 98.81	76.26 79.81 98.72	76.31 79.83 99.11	76.32 79.81 98.94	51.35 58.17 92.35	50.90 57.98 95.69	76.29 79.99 98.98	75.97 77.99 98.03
α β γ (°)	90 90 90											
Molecules per asymm. unit	2	2	2	2	2	2	2	2	1	1	2	2
R <sub>pim</sub>	0.125 (0.375)	0.067 (0.332)	0.047 (0.407)	0.043 (0.540)	0.029 (0.246)	0.029 (0.359)	0.029 (0.372)	0.046 (0.391)	0.088 (0.734)	0.101 (0.606)	0.048 (0.550)	0.028 (0.375)
CC <sub>1/2</sub>	0.974 (0.679)	0.994 (0.745)	0.997 (0.690)	0.998 (0.524)	0.998 (0.826)	0.999 (0.681)	0.999 (0.586)	0.998 (0.697)	0.996 (0.587)	0.990 (0.499)	0.997 (0.609)	0.999 (0.743)
I/σI	3.7 (1.0)	6.8 (1.0)	8.9 (1.0)	8.9 (1.1)	13.0 (2.3)	12.6 (1.3)	13.6 (1.4)	9.4 (1.4)	8.8 (1.1)	5.5 (1.0)	8.2 (1.1)	29.4 (3.5)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	99.8 (99.8)	99.0 (91.8)	100.0 (99.6)	99.8 (99.2)	100.0 (100.0)	100.0 (99.5)	99.5 (99.6)	99.7 (99.6)	99.8 (99.0)
Multiplicity	6.4 (6.5)	12.7 (9.3)	12.9 (10.9)	6.4 (6.4)	6.3 (4.7)	12.7 (10.0)	18.5 (10.4)	13.0 (13.5)	16.5 (6.3)	6.5 (6.1)	5.8 (5.9)	17.3 (13.0)
# reflections (total unique)	42219	51523	118990	113916	131151	121461	103204	80367	29625	27504	103164	114707
<b>Refinement</b>												
R work/free	0.1767/ 0.2164	0.1482/ 0.1863	0.1421/ 0.1665	0.1360/ 0.1665	0.1271/ 0.1412	0.1255/ 0.1409	0.1319/ 0.1480	0.1329/ 0.1518	0.1397/ 0.1615	0.1501/ 0.1794	0.1434/ 0.1633	0.1317/ 0.1497
<i>No. atoms</i>												
Protein	4754	5017	5261	5376	5661	5703	5407	5066	2398	2387	5164	5180
Ligand	–	24	–	24	–	24	–	24	–	12	–	24
Water	424	483	463	498	543	527	511	496	267	228	474	488
<i>RMSD</i>												
bond lengths (Å)	0.005	0.008	0.004	0.005	0.008	0.008	0.008	0.005	0.004	0.003	0.004	0.004
bond angles (°)	0.831	0.910	0.784	0.880	1.029	0.986	0.973	0.835	0.731	0.631	0.773	0.822
MolProbity clashscore	2.97	3.09	3.35	4.67	3.89	2.02	2.32	2.26	1.26	1.26	2.91	2.70

<sup>a</sup> Highest resolution shell is shown in parentheses.

**Supplementary Table 4.** Amino-acid positions optimized during computational design of HG4

Ligand placement <sup>a</sup>	Repacking <sup>b</sup>
16, 17, 21, 42, 44, 46, 47, 79, 81, 83, 84, 87, 90, 125, 130, 170, 172, 207, 209, 234, 236, 237, 239, 267, 275, 276	V16, Y17, A21, M42, W44, E46, N47, Q50, L79, G81, A82, G83, C84, W87, F90, T125, D127, G130, Y170, M172, Q207, H209, S234, L236, M237, D239, M267, A275, F276

<sup>a</sup> Positions that were mutated to Gly during ligand placement. Catalytic residues D127 and Q50 were allowed to sample alternate rotamers. All other residues were kept fixed.

<sup>b</sup> Positions and amino-acid types that were allowed to sample alternate rotamers during repacking. All other residues were kept fixed.

**Supplementary Table 5.** Geometric definitions for generation of transition-state poses off the side chains of catalytic residues

Contact	Type	Atom 1 <sup>a</sup>	Atom 2 <sup>a</sup>	Atom 3 <sup>a</sup>	Atom 4 <sup>a</sup>	Values <sup>b</sup>
<b>Asp127</b>	Distance	OD1 or OD2	<b>H3</b>			1.0, 1.2, 1.5
	Angle	CG	OD1 or OD2	<b>H3</b>		112, 117, 122
	Angle	OD1 or OD2	<b>H3</b>	<b>C3</b>		159, 164, 169, 174, 179
	Torsion	CB	CG	OD1 or OD2	<b>H3</b>	0, 5, 10, 170, 175, 180
	Torsion	CG	OD1 or OD2	H3	<b>C3</b>	170, 175, 180, 185, 190
	Torsion	OD1 or OD2	<b>H3</b>	<b>C3</b>	N2	0, 5, 170, 175, 180
<b>Gln50</b>	Distance	1HE2 or 2HE2	<b>O1</b>			1.2, 1.5, 1.7, 1.9, 2.1, 2.3
	Angle	NE2	1HE2 or 2HE2	<b>O1</b>		145, 148, 151, 154, 157
	Angle	2HE2	<b>O1</b>	<b>N2</b>		120, 125, 135, 145, 155
	Torsion	CD	NE2	1HE2 or 2HE2	<b>O1</b>	115, 120, 135, 140, 145
	Torsion	NE2	1HE2 or 2HE2	<b>O1</b>	N2	180, 190, 200, 210, 220, 230
	Torsion	1HE2 or 2HE2	<b>O1</b>	N2	<b>C3</b>	150, 160, 170, 180, 190, 200

<sup>a</sup> Atoms in bold are from the transition state. All other atoms are from the catalytic residues.

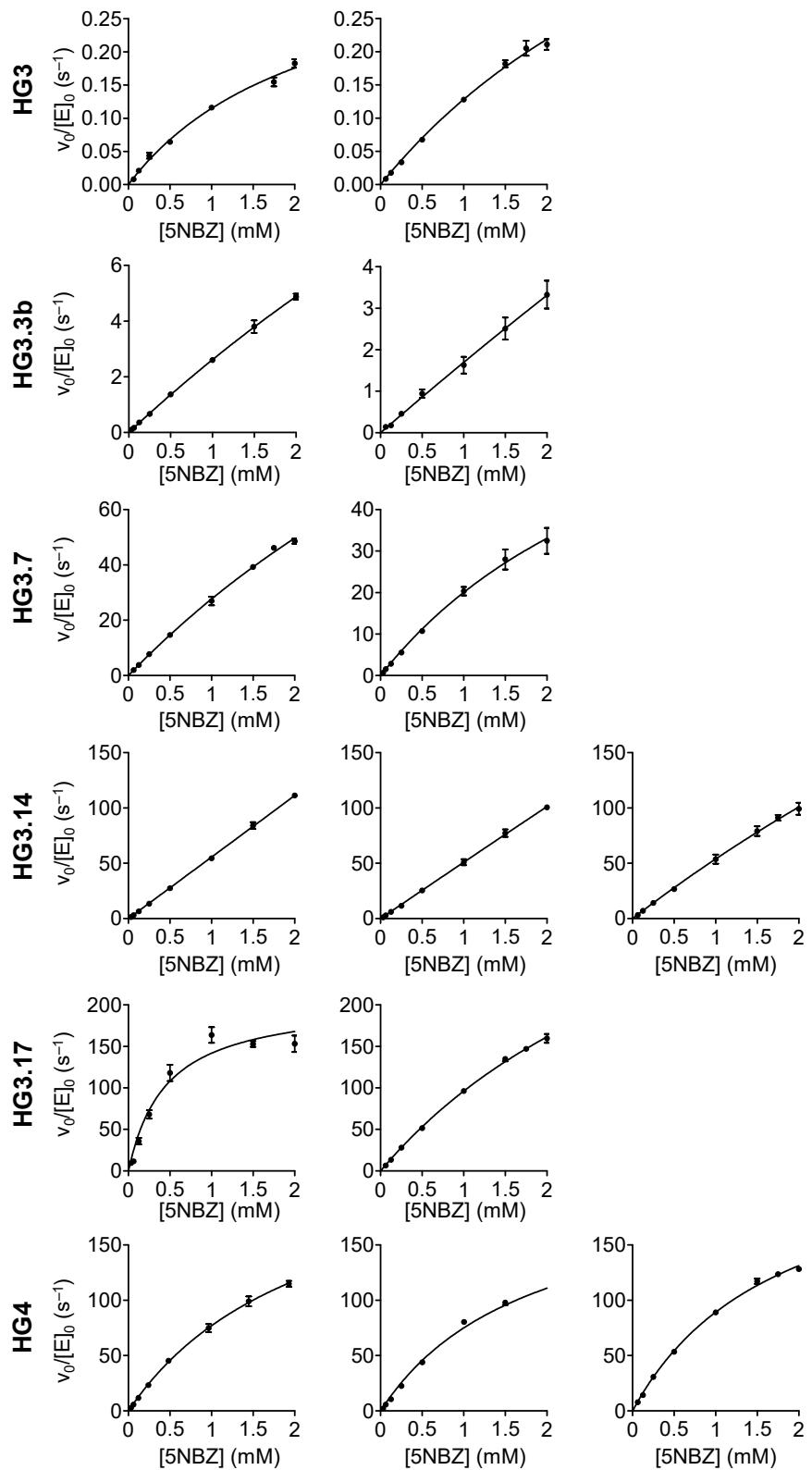
<sup>b</sup> Distance measurements given in Å, all others in degrees.

**Supplementary Table 6.** Geometric constraints used to define catalytic contacts during repacking step of HG4 computational design

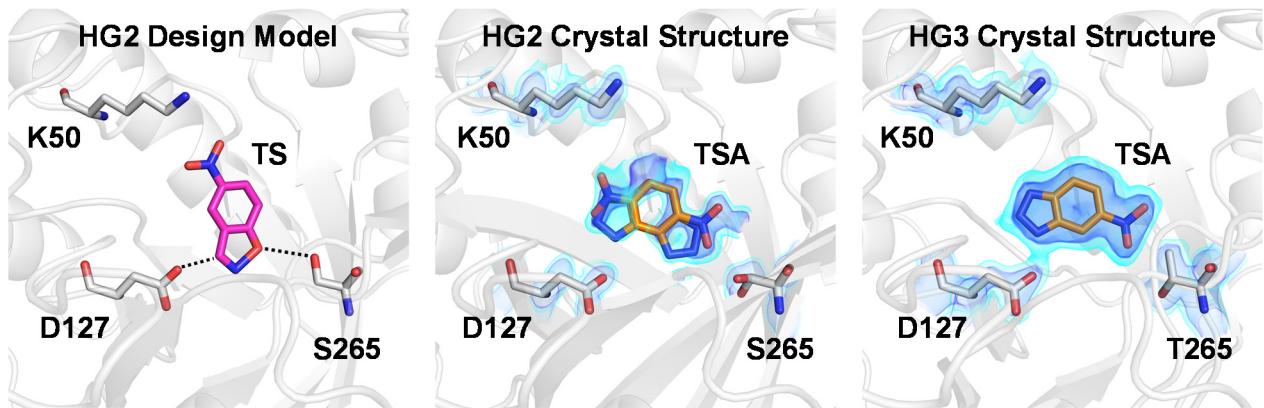
Contact	Type	Atom 1 <sup>a</sup>	Atom 2 <sup>a</sup>	Atom 3 <sup>a</sup>	Atom 4 <sup>a</sup>	Min <sup>b</sup>	Max <sup>b</sup>
<b>Asp127</b>	Distance	OD1 or OD2	<b>H3</b>			1.0 (1.0)	1.6 (1.6)
	Angle	CG	OD1 or OD2	<b>H3</b>		110 (110)	130 (130)
	Angle	OD1 or OD2	<b>H3</b>	<b>C3</b>		160 (160)	179 (179)
	Torsion	CB	CG	OD1 or OD2	<b>H3</b>	-20 (-20)	20 (20)
<b>Gln50</b>	Distance	1HE2 or 2HE2	<b>O1</b>			1.2 (1.2)	2.3 (3.2)
	Angle	NE2	1HE2 or 2HE2	<b>O1</b>		145 (131)	157 (179)
	Angle	1HE2 or 2HE2	<b>O1</b>	<b>N2</b>		119 (112)	139 (150)
	Torsion	1HE2 or 2HE2	<b>O1</b>	<b>N2</b>	<b>C3</b>	161 (129)	199 (199)

<sup>a</sup> Atoms in bold are from the transition state. All other atoms are from the catalytic residues.

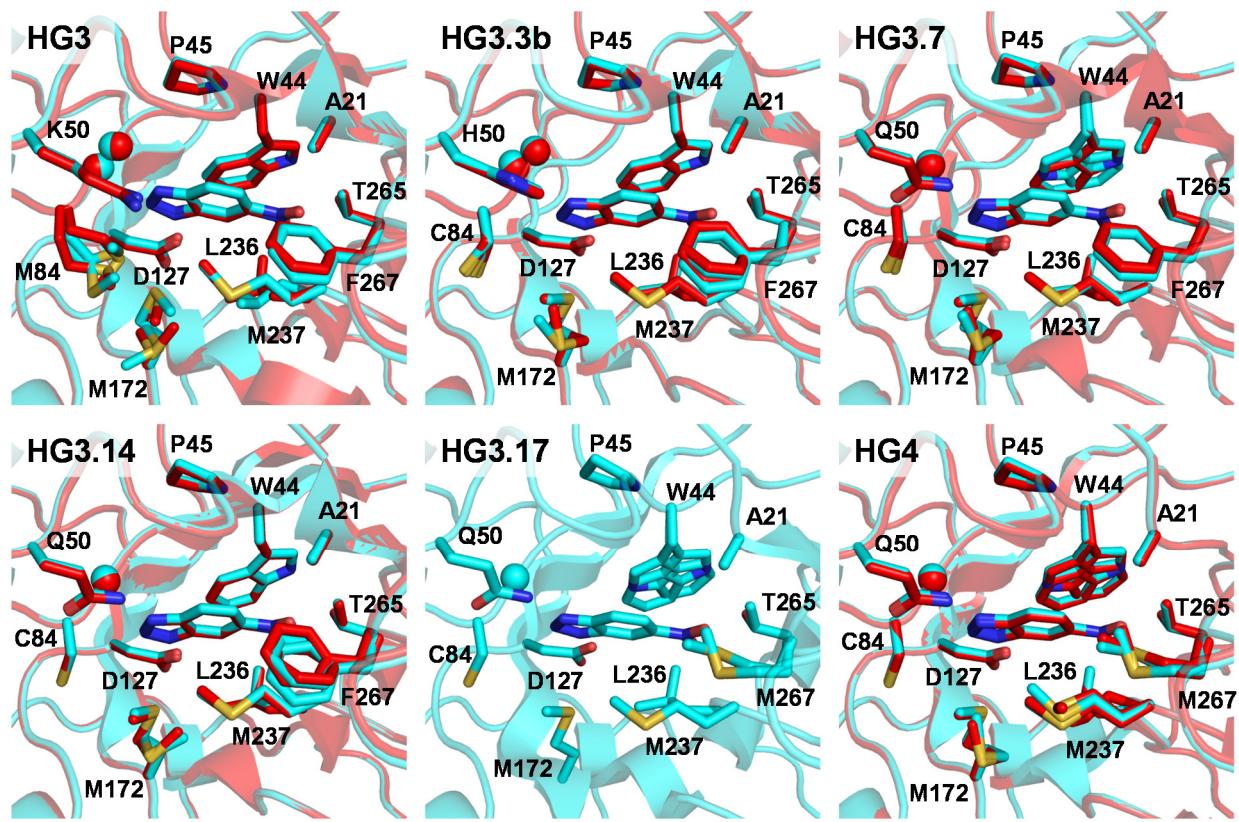
<sup>b</sup> Distance measurements given in Å, all others in degrees. Values in parentheses are for the repacking step, while the others are for ligand placement.



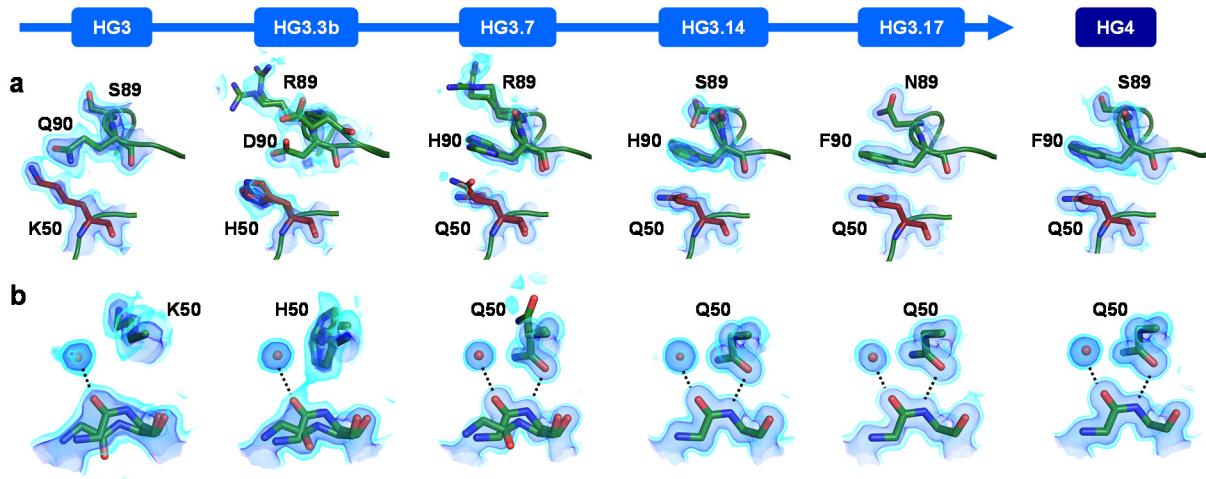
**Supplementary Figure 1. Steady-state kinetics.** Michaelis–Menten plots of normalized initial rates as a function of 5-nitrobenzisoxazole (5NBZ) concentrations are shown. All experiments were performed in triplicate (mean  $\pm$  s.d.) on at least two independent enzyme samples.



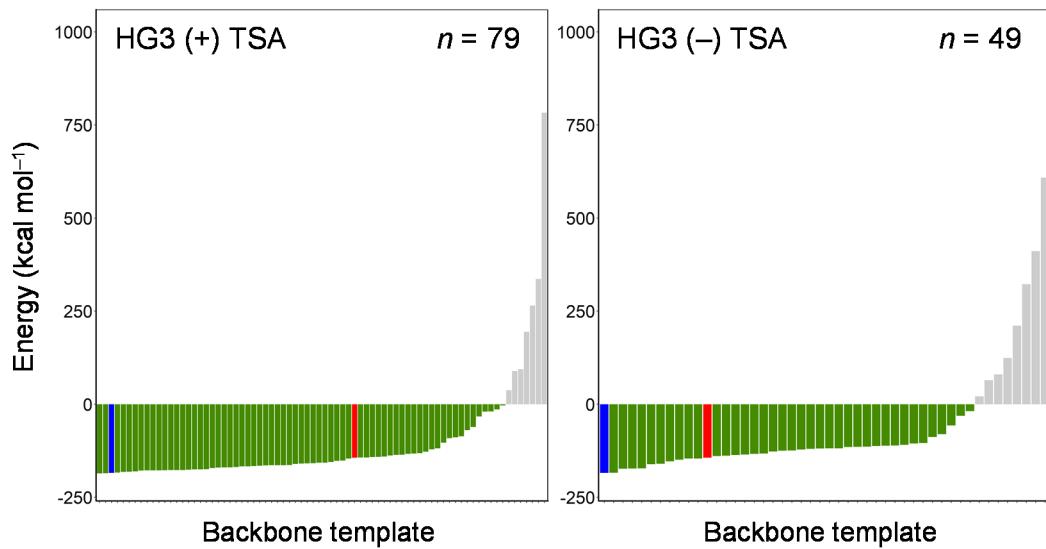
**Supplementary Figure 2. HG2 is the direct precursor to HG3.** For the crystal structures of HG2 and HG3, the 2Fo-Fc map is shown in volume representation at two contour levels:  $0.5 \text{ e}\text{\AA}^{-3}$  and  $1.5 \text{ e}\text{\AA}^{-3}$  in light and dark blue, respectively. HG2 was originally designed to stabilize the transition state (TS) via catalytic contacts (dashed lines) with the D127 base and the S265 hydrogen bond donor (Left panel). However, its crystal structure (PDB ID: 3NYD) showed that the transition state analogue (TSA) was bound in two alternate orientations (Middle panel). In the catalytically productive pose, the acidic N-H bond of the TSA that mimics the cleavable C-H bond of the substrate is located within H-bonding distance to the carboxylate oxygen of D127, and the nitro group is close to S265. In the catalytically non-productive pose, the TSA is flipped, which positions its nitro group closer to K50, and its acidic N-H bond far from the side chain of D127. To increase activity, Privett *et al.* introduced the S265T mutation into HG2, leading to HG3 (Right panel). This mutation was predicted by molecular dynamics to rigidify the active site, and thereby increase activity. Of note, no density for the non-productive binding pose of the TSA was observed in the HG3 crystal structure reported here.



**Supplementary Figure 3. TSA-binding pocket.** Overlay of chains A (cyan) and B (red) showing residues forming the TSA-binding pocket in all HG-series Kemp eliminases. For HG3.17, only chain A is shown as its asymmetric unit contained a single polypeptide chain. In all cases, the TSA is bound in the middle of the barrel. Spheres indicate alpha carbons of Gly83 (two conformers, *cis/trans*, are observed for peptide bond between Gly83 and Met/Cys84 in HG3 and HG3.3b).



**Supplementary Figure 4. Crystal structures of HG-series Kemp eliminases in the unbound state.** In all cases, only atoms from chain A are shown. The 2Fo-Fc maps are shown in volume representation at two contour levels: 0.5 eÅ<sup>-3</sup> and 1.5 eÅ<sup>-3</sup> in light and dark blue, respectively. (a) Conformational changes to loop formed by residues 87–90 over the course of the evolutionary trajectory. (b) The peptide bond between residues 83 and 84 adopts both *cis* and *trans* conformations in HG3, HG3.3b, and HG3.7, but only the *cis* conformation in the higher activity variants. Ordered water molecules are shown as red spheres, and hydrogen bonds as dashed lines.



**Supplementary Figure 5. Energy of HG4 design models generated on various backbone templates.** Rotamers for the HG4 sequence and its associated transition state binding pose were optimized (Methods) on individual backbone templates (bars). These ensembles of backbone templates were generated using molecular dynamics fitted to the HG3 diffraction data with and without TSA. In all cases,  $n$  indicates the number of templates per ensemble. Green and grey bars indicate templates that yielded design models with favorable or unfavorable energy, respectively. Blue and red bars indicate design models obtained from the HG4 with bound TSA ( $-184.4$  kcal/mol) or 1GOR ( $-143.5$  kcal/mol) crystal structures, respectively. Several templates from each ensemble yielded HG4 design models with more favorable energy than that obtained on the 1GOR template, but only the HG3 with TSA ensemble yielded models that scored more favorably than the one obtained on the HG4 (+) TSA crystal structure.