nature chemistry

Article

https://doi.org/10.1038/s41557-026-02125-6

Emergence of specific binding and catalysis from a designed generalist binding protein

Received: 18 July 2025

Accepted: 6 March 2026

Published online: 04 May 2026

Check for updates

Yuda Chen[1,2,11], Sagar Bhattacharya[1,2,11], Lena Bergmann[3,11], Galen J. Correy[3,11], Sophia K. Tan[1,2], Kaipeng Hou[1,2], Justin T. Biel[3], Lei Lu[1,2], Ian Bakanas[1,2], Jason E. Gestwicki[1,4], Alexander N. Volkov[5,6], Ivan V. Korendovych[7], Nicholas F. Polizzi[8,9], James S. Fraser[3] ✉ & William F. DeGrado[1,2,10] ✉

Binding and catalysis play central roles in living systems. While natural proteins have finely tuned affinities for their primary ligands, they also bind weakly and promiscuously to other molecules, which serve as starting points for the incremental evolution of different specificities. Thus, modern proteins have emerged from the joint exploration of sequence and structural space. Interactions between natural proteins and small molecules can be systematically profiled by crystallographic fragment screening in defined geometries, yet this approach has not been applied to highly designable de novo proteins. Here we apply this method to explore the binding specificity of a de novo small-molecule-binding protein, apixaban-binding helical bundle. As in nature, we found that it formed weak complexes, which were excellent starting points for the design of entirely distinct functions, including a turn-on fluorophore binder and a highly efficient Kemp eliminase with a catalytic efficiency of 3,200,000 $M^{-1} s^{-1}$, approaching the diffusion limit. This work illustrates how simultaneous consideration of sequence and chemical structure diversity can guide the emergence of different functions in designed proteins.

Proteins form the essential machinery of life. Understanding their functions has practical implications for the design and engineering of catalysts and pharmaceuticals. Previously, Tawfik and co-workers demonstrated a nuanced mechanism underlying the ongoing evolution of distinct functions in proteins[1,2]. Highly evolved proteins are generally specific to their substrates, yet their active sites often display a degree of 'promiscuity', binding various molecules with specific geometries and functions, but with lower affinity than for the native ligand[1]. Such weak interactions can serve as starting points for the emergence of different functions by an accrual of mutations on an evolutionary timescale[2] (Fig. 1a). In this mechanism, a 'specialist' protein for binding a biologically important ligand can also serve as a 'generalist' for binding weakly to a diverse set of molecules. Given a positive genetic selection, the generalist activity can serve as a starting point for the evolution of a distinct specialist protein with increased affinity or catalytic activity for one member of the initial set of compounds that was initially bound with only low affinity. Here, inspired by this concept, we have computationally designed a

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA. [2]Cardiovascular Research Institute, University of California, San Francisco, CA, USA. [3]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA. [4]Institute for Neurodegenerative Diseases, University of California, San Francisco, CA, USA. [5]VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie (VIB), Brussels, Belgium. [6]Jean Jeener NMR Centre, Vrije Universiteit Brussel (VUB), Brussels, Belgium. [7]Department of Chemistry and Biochemistry, Baylor University, Waco, TX, USA. [8]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [9]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. [10]Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai, Japan. [11]These authors contributed equally: Yuda Chen, Sagar Bhattacharya, Lena Bergmann, Galen J. Correy. ✉e-mail: jfraser@fraserlab.com; bill.degrado@ucsf.edu

https://doi.org/10.1038/s41557-026-02125-6

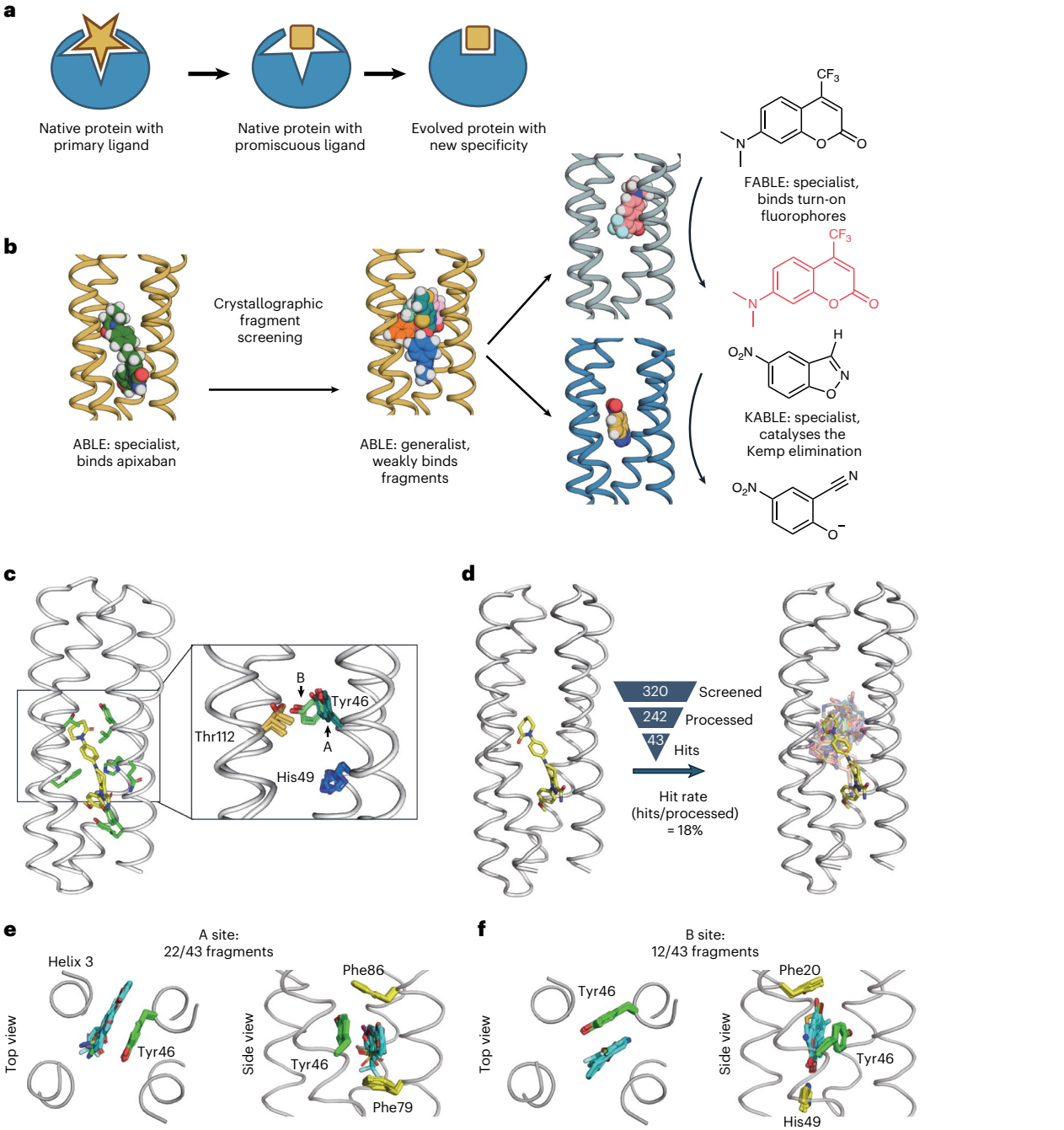


Fig. 1 | Nature-inspired evolution of de novo proteins. a, A natural protein that has specialized for the binding or catalysis of one molecule (star) can also exhibit promiscuous and weak binding to a second molecule (square). The protein can alter the shape of its binding site to match the second molecule by shifts in conformation and changes to the residues lining the binding pocket. **b**, Here we use crystallographic fragment screening to probe the ability of a de novo apixaban-binding protein, ABLE, to bind small molecules promiscuously with low affinity. We then redesign the protein to bind turn-on fluorophores and catalyse the Kemp elimination reaction. **c**, Multiple side-chain conformations observed at the active site residues of different ABLE crystal structures. The apixaban-bound ABLE structure (Protein Data Bank (PDB) 6W70) is shown on the left with its binding site residues shown as green sticks. Inset: multiple side-chain conformations of Tyr46, His49 and Thr112 in apixaban-bound ABLE (PDB 6W70) and the ligand-free forms (PDB 6W6X and 6X8N). These crystal structures contain multiple monomers in their asymmetric units, each of which is superimposed onto the backbone of chain A of apixaban-bound ABLE in the inset. Two distinct groups of Tyr46 conformations were observed. The group containing the Tyr46 conformation from the apixaban-bound ABLE structure (PDB 6W70) was labelled as A, and the other group was labelled as B. **d**, Crystallographic fragment screening reveals the chemical space of weak binders. The crystal structures of ABLE complexed with 43 fragments were each solved to between 1.3 and 1.6 Å resolution using Pan-Dataset Density Analysis[21]. The structures of the complexes superposed on the ABLE complex (PDB 6W70, chain A) are shown with the carbon atoms of apixaban as yellow sticks, and the carbon atoms of the remaining fragments shown in different colours. The hit rate was calculated by dividing the number of fragment hits by the number of processed datasets. **e**,**f**, Most fragments interact with ABLE at one of two binding sites, designated A (**e**) and B (**f**), which differ by a conformational shift of Tyr46. Two views of the complexes are shown looking down the axis of the bundle (left) and rotated by approximately 90° (right). The active sites of multiple ABLE–fragment complexes, categorized as either the A or B site, are shown with the interacting residues as green sticks and the fragments as blue sticks. The A and B sites were named based on the inclusion of Tyr conformers A and B, respectively, which was explained at **c**. Only one backbone (7HK4 for **e** and 7HK3 for **f**) is shown because of highly conserved backbone structures.

previously de novo-designed ligand-binding protein to have distinctive ligand-binding and catalytic activities.

The development of de novo protein design[3–6] has been influenced by these hypothetical mechanisms of protein evolution. However, functional elaboration of de novo protein scaffolds has the advantage that an arbitrary number of amino acid insertions, deletions and substitutions can be simultaneously made through design in ways that are not feasible through the more incremental stepwise process of natural protein evolution[7–11]. Indeed, using computational design, de novo proteins have been designed for various functions[7–11]. However, in each case, the design process has been directed towards a single, specialized function. While the specificity of the proteins for closely related ligands[12] or substrates[13–18] has been evaluated in isolated cases, the large-scale structural examination of the specificity of de novo proteins has not been reported. Recent advances in X-ray crystallography throughput now enable the detection of low-affinity molecules bound to receptor proteins[19]. In drug discovery, this is applied by soaking hundreds to thousands of small-molecule fragments (<300 Da) into crystals and identifying binders through changes in the electron density[20]. Real-space background subtraction enhances sensitivity, allowing the detection and modelling of bound molecules with typical hit rates of 1–20% (ref. 21). In this study, we determined the specificity of a de novo protein towards common chemotypes in a diverse set of small molecules using crystallographic fragment screening and then used this information to guide the design of entirely distinct binding and catalytic functions.

We focused on apixaban-binding helical bundle (ABLE), a de novo protein that was designed to bind with high specificity to the anti-thrombotic drug, apixaban[22]. Inspired by Tawfik and co-workers' hypothesis[23], we examined how ABLE, initially designed as a specialist binder of apixaban, could give rise to distinct specialists: one fluorophore binder with high specificity for a different small molecule, and a catalyst whose activity exceeds that of many proficient natural enzymes[24–27] (Fig. 1b). We initially explored ABLE's promiscuity by screening a library of several hundred small-molecule fragments at millimolar concentrations using X-ray crystallography. Through this approach, weak molecular interactions were identified that were not anticipated in the initial designs, which provided pathways to intended specificities and catalytic activity, achieved through computational design.

In this study, we used helical bundles, which are one of the most ancient and adaptable protein architectures. Most transmembrane proteins, including G-protein coupled receptors (GPCRs), rely on helical bundles for signal transduction, transport and enzymatic functions[28]. The enzymes that catalyse some of the most important reactions for the evolution of life, such as the formation and utilization of O_2 and the generation of ion and proton gradients, are also helical bundles[29–31]. Similarly, soluble helical bundle proteins are involved in electron transport[32], small-molecule binding[33] and catalysis[34]. By linking modern protein design efforts to the fundamental principles of catalysis, molecular recognition and evolution, we have designed distinct specificities and catalytic functions starting with a common structural framework.

Results

Conformational plasticity and binding specificity of ABLE from crystallographic fragment screening

While de novo proteins have been designed to discriminate variants of their substrates, their ability to bind with low affinity to structurally unrelated substrates has not been explored. We therefore evaluated the promiscuity of ABLE, which was designed to bind the anti-thrombotic drug apixaban in an extended groove within a helical protein (Fig. 1c).

We used X-ray crystallography to map regions of ABLE's binding site that are capable of binding to a diverse set of 320 'fragments' (molecular weight < 300 Da) of drug-like molecules with a wide range of aliphatic, aromatic and polar functional groups[35]. Each compound (10 mM) was co-incubated with ABLE crystals to form complexes in the crystal lattice. We obtained high-quality diffraction data from 242 of the compounds in the library. Of these, 43 gave diffraction data with electron density that could be ascribed to a bound fragment (Fig. 1d).

The frequency of forming solvable complexes provides a measure of the promiscuity of the site and the structures of the resulting complexes provide a view of the diversity of interactions that can be formed within the binding site. The hit rate for discovering fragments bound to ABLE (18%) is similar to the reported hit rates of 7% and 19% for the Nsp3 macrodomain of the SARS-CoV-2 and Chikungunya viruses, respectively, using the same fragment library and screening method[20,36] (Fig. 1d). All but four fragments bound to ABLE within the apixaban binding site. Thus, by this criterion, ABLE has a specificity within the range seen in natural proteins.

The 39 fragments that bound in ABLE's apixaban-binding site define a set of chemotypes that favourably interact with ABLE (Extended Data Figs. 1–5). The bound fragments have a notable tendency to be slightly larger, more apolar and richer in aromatic rings when compared with the overall library (Supplementary Fig. 1). The preferred interacting groups found in the complexes show some differences from the interactions that are important for binding the drug apixaban. For example, while apixaban has a primary carboxamide (-$CONH_2$) that is bound near the surface of ABLE in the drug complex, this polar group was not seen in the bound fragments, all of which bound more deeply in the pocket.

The protein's structure was nearly identical in all of the complexes (sub-ångström Cα root mean squared deviation (RMSD)). About two-thirds of the fragments bound to A site, where Tyr46 adopts conformation A, which is found in the apixaban complex, featuring an aromatic box comprising the side chains of Tyr46, Phe79 and Phe86 and the backbone of helix 3 (Fig. 1e). However, a considerable fraction (12/43) of the fragments bound to B site, where a distinct Tyr46 conformer B occurs and Tyr46, Phe20 and His49 form another aromatic box (Fig. 1f). A modest difference in the torsional angles of Tyr46's side chain in conformers A and B propagates a larger displacement of its long aryl side chain. In addition to aromatic and apolar interactions, most fragments also form polar interactions. For example, Thr112 often engages polar groups in conformer A, as does His49 in conformer B (Extended Data Figs. 2–5 and Supplementary Fig. 2).

Often, generalist proteins bind additional substrates in conformational sub-states that become locked in place as the protein evolves specificity for the additional substrates[1]. We were therefore curious whether some of the additional conformers seen in the fragment complexes, but not in the apixaban complex, might pre-exist in the multiple crystal structures of unliganded ABLE that we have solved. Indeed, we observed that the critical Tyr46 residue adopted both conformers A and B in different structures of unliganded ABLE (Fig. 1c). A similar pair of conformations emerged from multiple molecular dynamics simulations of the protein in the drug-free state (Extended Data Fig. 6). Other sources of conformational variability in this region include the side chains of His49, which also populate more than one rotamer (Supplementary Fig. 3). Thus, key elements of conformer B are already found in the uncomplexed protein, indicating that ABLE does not pay a large reorganization penalty to adopt conformer B.

Designing proteins to bind fluorogenic ligands

We next asked whether conformer B would be a good starting point for the design of specialized proteins with altered binding specificity. Site B bound the fragment 7-hydroxycoumarin (Fig. 2a), which prompted us to redesign ABLE to bind the larger, brighter and environment-sensitive turn-on fluorophore 7-(dimethylamino)-4-(trifluoromethyl)coumarin (Cou485; Fig. 2a)[37]. Procedurally, this task involved expanding the top of the B site to accommodate the additional trifluoromethyl and dimethylamino groups while filling the lower, unoccupied section of

https://doi.org/10.1038/s41557-026-02125-6

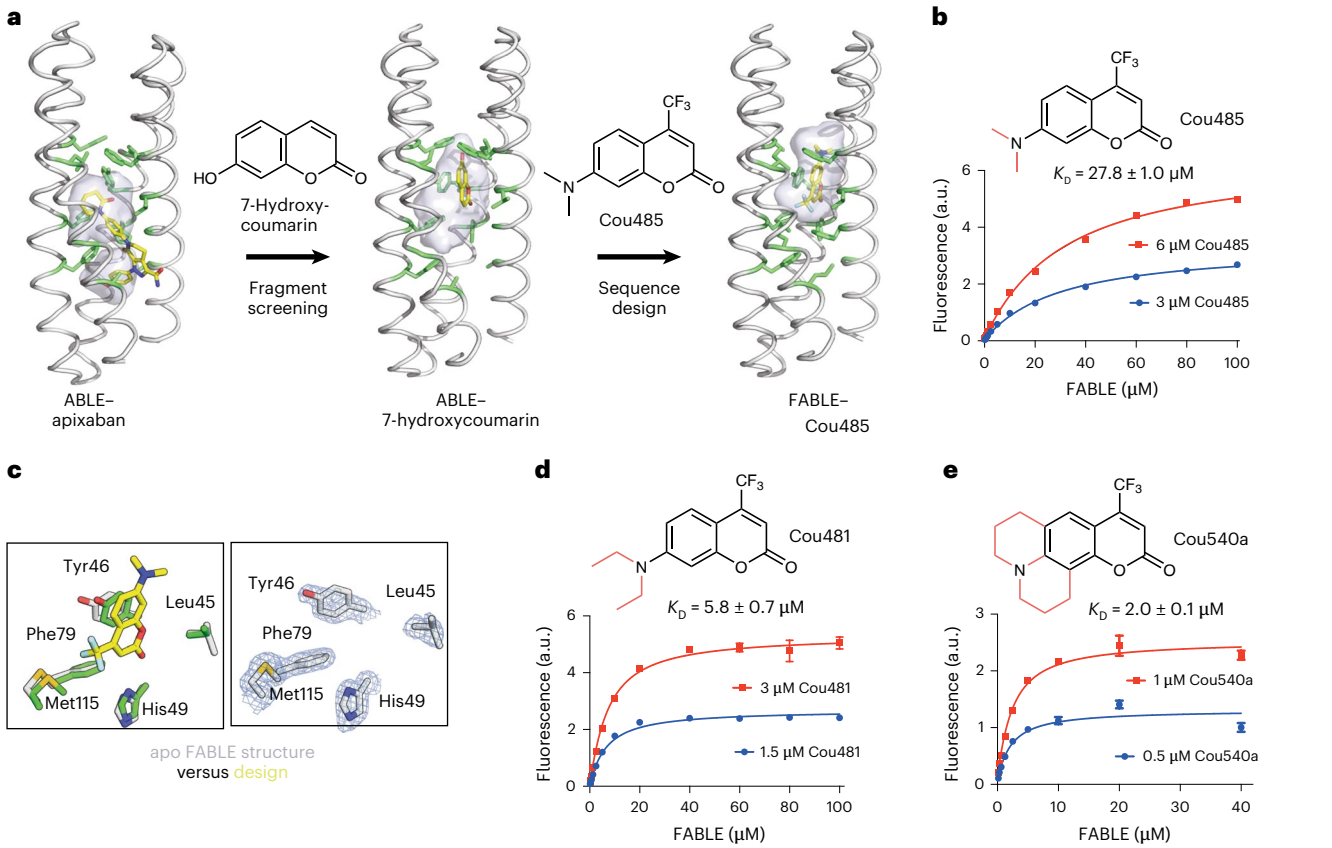


Fig. 2 | Design of FABLE binding to fluorogenic coumarins. a, 7-Hydroxycoumarin was found to bind ABLE. The Tyr46 B conformer (Fig. 1f) was used to design five sequences for binding to Cou485. The apixaban-interacting residues (green sticks) and the binding cavity (grey surface) show how the site transitions from binding apixaban, then a small coumarin and then Cou485. **b**, Titration of FABLE into Cou485 shows a single-site binding isotherm. The dissociation constant was obtained by globally fitting a single-site binding model to the data obtained by titrating FABLE into Cou485 at two fixed fluorophore concentrations of 3 and 6 μM. **c**, Left: overlay of the designed FABLE–Cou485 complex with the uncomplexed FABLE crystal structure (PDB 9DWC; the comparison with 9DWA and 9DWB is shown in Supplementary Fig. 8). Right: $2mF_o - DF_c$ omit electron density (1σ contour) of the apo FABLE crystal structure (PDB 9DWC). The design structure shows Cou485 as yellow sticks, with the side chains of interacting residues shown as green sticks. The corresponding residues in the apo FABLE crystal structure are shown as grey sticks. **d**,**e**, The binding constants for FABLE's interaction with Cou481 (**d**) and Cou540a (**e**) were determined for fixed fluorophore concentrations titrated with variable FABLE concentrations as in **b**. These two compounds were identified after screening a set of 37 fluorophores, including coumarin derivatives and other fluorophore scaffolds. The error bars in **b**, **d** and **e** represent standard deviations from three independent measurements. Error bars in **b** are smaller than the symbol size and are therefore not visible.

the pocket with well-packed hydrophobic amino acids (Fig. 2a). Using a previously described protocol[38,39], we selected 43 of ABLE's 126 residues to vary during redesign. Figure 2a illustrates the gradual migration of the binding site as it binds the larger ligand. Structure prediction using OmegaFold[40] indicated that the designs would be preorganized in a conformation capable of binding Cou485; other selection metrics are described in Methods.

Five of the designed proteins with sequence pairwise identity ranging from 77% to 81% were selected for experimental characterization (Supplementary Fig. 4). All five were helical, thermostable to 90 °C and bound Cou485 (Extended Data Fig. 7). The fluorescence of Cou485 (6 μM) increased 10- to 99-fold in the presence of 40 μM of the five different proteins (Supplementary Fig. 5), indicative of binding in a rigid, hydrophobic binding site[41]. Global analysis of the titration curves obtained for two fixed Cou485 concentrations and variable protein concentrations indicated that all five proteins bound Cou485 with dissociation constants (K_D) <100 μM, and three exhibited K_D ranging from 28 to 38 μM (Extended Data Fig. 7). The protein showing the most potent K_D (27.8 ± 1.0 μM) was designated as fluorescent ABLE (FABLE; Fig. 2b). Attesting to their specificity, the starting ABLE protein did not show a saturable curve, indicative of very weak binding (Supplementary Figs. 5 and 6).

Ligand efficiency is an empirical metric used to assess the effectiveness of a protein's interaction with a small-molecule ligand[42,43]. Large ligands have more opportunities to interact favourably, so ligand efficiency roughly normalizes to the size of the molecule by dividing the free energy of binding (1 M standard state) by the number of heavy atoms in the ligand. Most small-molecule drugs approved by the Food and Drug Administration have a ligand efficiency of around 0.3 kcal per mol per heavy atom count[43]. By comparison, the ligand efficiencies for the five proteins with Cou485 range from 0.31 to 0.35 kcal per mol per heavy atom count (Supplementary Table 1). To address the question of whether FABLE has a well-defined, preorganized structure, we determined crystallographic structures of ligand-free FABLE in three different crystal forms (1.5–1.9 Å resolution). Each of the structures was in excellent agreement (<1 Å Cα RMSD) with the designed model and the structure predicted by AlphaFold 3 (Fig. 2c, Supplementary Figs. 7 and 8, and Supplementary Table 2).

Experimental exploration of chemical space to probe the specificity of FABLE

Having explored the sequence space to create a specialist that is specific for Cou485, we next re-examined the chemical space to evaluate FABLE's specificity for alternative fluorophores. A search of 37 fluorophores

https://doi.org/10.1038/s41557-026-02125-6

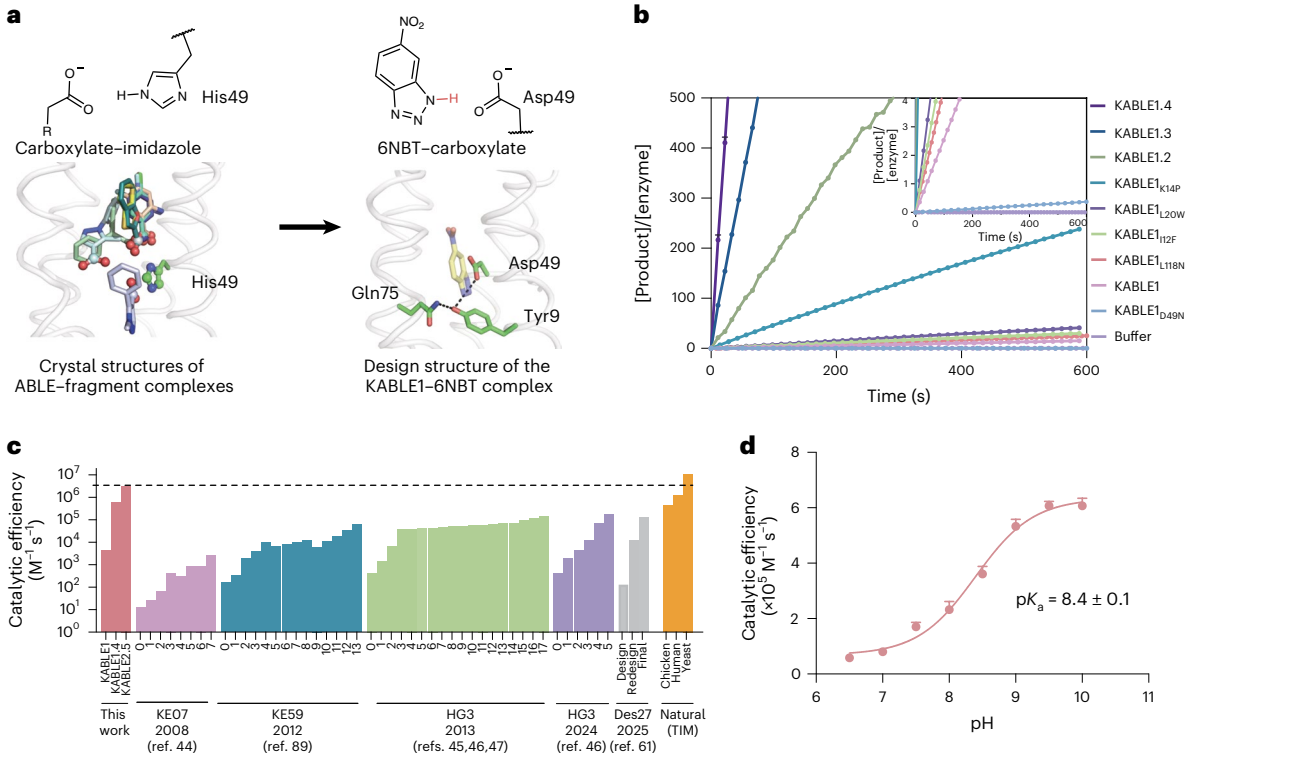


Fig. 3 | Design of an efficient KABLE and its directed evolution. a, Fragment-inspired design of KABLE. The carboxylate–imidazole interaction is analogous to the carboxylate–isoxazole interaction at the active site of a Kemp eliminase, suggesting that an active catalyst might be designed by changing His49 to an Asp. Left: nine of the fragments contain carboxylates that bind to His49. Right: design structure of KABLE1–6NBT. **b**, Catalytic turnover for the Kemp elimination reaction catalysed by KABLE1 (the original designed sequence) and its mutants. The mutants represent the single beneficial mutations from site-saturation mutagenesis and their combinations. The reactions were performed at pH 8.0 with an initial substrate concentration of 240 μM. **c**, Catalytic efficiency of the designed enzyme and its optimized variants, obtained through directed evolution or computational redesign, compared with triosephosphate isomerase (TIM), a natural enzyme with a similar mechanism. For each protein, each step involved in the iterative improvement is indicated. The first on the left reflects the activity of the initial design; the subsequent bars reflect the activities after saturation mutagenesis (KABLE, this work), combining favourable mutants after multiple rounds of directed evolution (KE07, KE59 and HG3)[44–47,89] or computational redesign[61]. The data for three homologues of TIM from different species are shown[25–27]. The dashed line indicate the catalytic efficiency of KABLE2.5. **d**, pH dependence of k_{cat}/K_M for KABLE1.4. k_{cat}/K_M values obtained at different pH were used to fit $(k_{cat}/K_M)_{max}$ and pK_a. The error bars in **d** represent standard errors of the mean from three independent measurements.

that shared molecular features such as hydrophobicity and aromaticity (Supplementary Table 3) identified only two compounds that showed a large (more than fourfold) increase in fluorescence intensity in the presence of FABLE compared with the starting ABLE protein (Extended Data Fig. 8). These compounds, Cou481 and Cou540a, are derivatives of Cou485 with slightly larger apolar groups at the 7-amino position of the coumarin core, which bound FABLE with $K_D = 5.8 \pm 0.7$ and 2.0 ± 0.1 μM (Fig. 2d,e and Supplementary Fig. 9) and ligand efficiencies of 0.35 and 0.34, respectively (Supplementary Table 1). The maximal fluorescence enhancement at saturating concentrations of FABLE was 48-fold for Cou485, 123-fold for Cou481 and 117-fold for Cou540a (Supplementary Fig. 9). Thus, the generalist binding activity of ABLE provides an excellent starting place for a specialist that binds a chemically distinct compound.

Design of an efficient Kemp eliminase

We next asked whether the generalist binding activity of ABLE could be used to engineer enzymatic function. We noted that many of the fragments that bound to the aromatic boxes are similar in shape and properties to the substrate used in Kemp elimination, a reaction that has been benchmarked over several decades in the quest to design enzyme-like proteins[44–53]. This reaction involves the removal of a proton from a C–H bond of the substrate (5-nitrobenzisoxazole), similar to the carboxylate-mediated abstraction of a proton from a C–H bond in the substrates of triosephosphate isomerase and ketosteroid isomerase[25,45,54]. The success rates and catalytic efficiencies of computationally designed Kemp eliminases were relatively low before rational redesign or directed evolution (0–200 M^{-1} s^{-1})[44,47,48]. Previous computationally designed Kemp eliminases were based on the redesign of the active sites of natural enzymes. We were curious whether a greater rate might be achieved with a de novo scaffold and our knowledge of ABLE's potential binding interactions.

To design a series of Kemp eliminases based on ABLE (KABLE proteins), we targeted a transition state analogue (TSA), 6-nitrobenzotriazole (6NBT; Fig. 3a). We computationally searched[47,48] for backbone positions of ABLE in which a Glu or Asp in a low-energy rotamer could position 6NBT in a catalytically competent orientation in the pocket delineated in fragment complexes. L108E was one site chosen because a low-energy rotamer of Glu108 placed the TSA inside the aromatic box of the A conformer. We also evaluated His49 as an attractive position for substitution of an Asp residue because histidine has a similar size, shape and polarity to aspartic acid. Importantly, His49 is located near the bottom of the binding site, which is a hotspot for interactions with carboxylate-containing fragments (Fig. 3a and Supplementary Fig. 10). Modelling an Asp side chain at residue 49 reverses the interactions and allows a strong hydrogen bond between 6NBT and the more basic (*syn*) lone-pair electrons of the Asp carboxylate[55] (Fig. 3a). With these restraints in place, we used LigandMPNN[56] and Rosetta FastRelax[39] to optimize the sequence to bind 6NBT while stabilizing the conformation of either Glu108 or Asp49. To increase their basicity, we introduced no more than one hydrogen-bond donor to the Asp49 or Glu108 carboxylate, leaving three potential

https://doi.org/10.1038/s41557-026-02125-6

hydrogen-bond acceptor sites unsatisfied[57]. Computed models were filtered according to the RMSD between the design model and that predicted for the unliganded protein computed by RaptorX[58] and ESM2[59] (to assess the preorganization of the substrate-free protein), as well as the metrics (Supplementary Fig. 11) that assess the likelihood of the predicted structures.

We expressed and purified five proteins from each strategy for experimental characterization, with pairwise sequence identity ranging from 26% to 71% compared with ABLE (Supplementary Fig. 4). One of the sequences (designated KABLE0) from the first strategy showed modest catalytic activity (k_{cat}/K_M = 8 M^{-1} s^{-1} at pH 8.0; Supplementary Fig. 12). A second protein (designated KABLE1) with H49D showed pH-dependent activity (apparent pK_a = 8.8 ± 0.6) with a maximal k_{cat}/K_M value of 3,800 ± 1,400 M^{-1} s^{-1} (k_{cat}/K_M = 460 ± 20 M^{-1} s^{-1} at pH 8.0; Fig. 3b,c, Extended Data Tables 1 and 2, Supplementary Tables 4–6 and Supplementary Fig. 13). Replacing Asp49 with Asn eliminated the catalytic activity, indicating that this residue is likely to be the protonatable group (Fig. 3b and Supplementary Fig. 14). The elevated pK_a for Asp49 is consistent with its placement in a very apolar environment, as seen in a Kemp eliminase based on ketosteroid isomerase[52,60]. Asp49 is a part of a hydrogen-bonded and essential 'Asp triad' network consisting of Tyr9 and Gln75 (Fig. 3a). Replacement of either Tyr9 or Gln75 also greatly decreased the eliminase activity of KABLE1 (Supplementary Fig. 14).

We next increased the catalytic efficiency 170-fold through saturation mutagenesis (Fig. 3b–d and Extended Data Table 1). We identified four single mutants (I12F, K14P, L20W and L118N), which individually increased activity 1.3- to 8.3-fold at pH 8.0 (Fig. 3b and Supplementary Table 5), and were also functionally complementary when combined (Fig. 3b, Supplementary Table 5 and Extended Data Table 1). Thus, the catalytic efficiency increased by over two orders of magnitude in the quadruple mutant, KABLE1.4, with a k_{cat}/K_M of 230,000 ± 30,000 M^{-1} s^{-1} at pH 8.0 (Fig. 3d, Extended Data Table 1). Its pH dependence revealed a maximal value of k_{cat}/K_M of 630,000 ± 20,000 M^{-1} s^{-1} and an apparent pK_a of 8.4 ± 0.1 (Fig. 3d and Extended Data Table 1).

We further improved the activity of KABLE1.4 by introducing substitutions at positions outside of the first shell of contacts using nuclear magnetic resonance (NMR)-guided directed evolution[49]. The amide proton resonances of a protein are particularly sensitive to small changes in local structures and can be used to identify residues whose environments are affected by the addition of a transition state analogue. The addition of 6NBT to KABLE1.4 induced large changes in its ^1H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum (Fig. 4a) for 24 residues (Supplementary Fig. 15). Saturation mutagenesis at these sites identified five positions at which single-residue substitutions increased activity 1.2- to 2.5-fold (Fig. 4b and Supplementary Table 5). Combining these mutants resulted in KABLE2.5, which showed a fivefold improvement in k_{cat}/K_M at pH 8.0 relative to KABLE1.4 (Supplementary Table 5). Replacing the Asp49 active site of KABLE2.5 with an Asn abolished activity, and substitutions of the other members of the catalytic triad (Y9F and Q75M) decreased k_{cat}/K_M more than tenfold (Supplementary Fig. 16).

While our initial characterization was conducted using a plate reader, we turned to stopped-flow ultraviolet–visible (UV–vis) spectroscopic measurements to obtain high-quality kinetic data. The value of k_{cat}/K_M for KABLE2.5 at pH 8.0 thus determined was 2,500,000 ± 180,000 M^{-1} s^{-1} (k_{cat} = 550 ± 10 s^{-1} and K_M = 0.210 ± 0.01 mM; Fig. 4c,d and Supplementary Fig. 17), which is within experimental error of the value obtained using a conventional cuvette and manual mixing (Supplementary Table 5). The stopped-flow method facilitated accurate measurements of the pH dependence of KABLE2.5, which conforms closely to a single-site isotherm with an apparent pK_a of 7.3 ± 0.1 (Fig. 4c and Extended Data Table 1).

We compared KABLE2.5 with HG3.17 (ref. 45), an efficient Kemp eliminase that was computationally designed starting from a natural enzyme[47] and improved through 17 rounds of directed evolution.

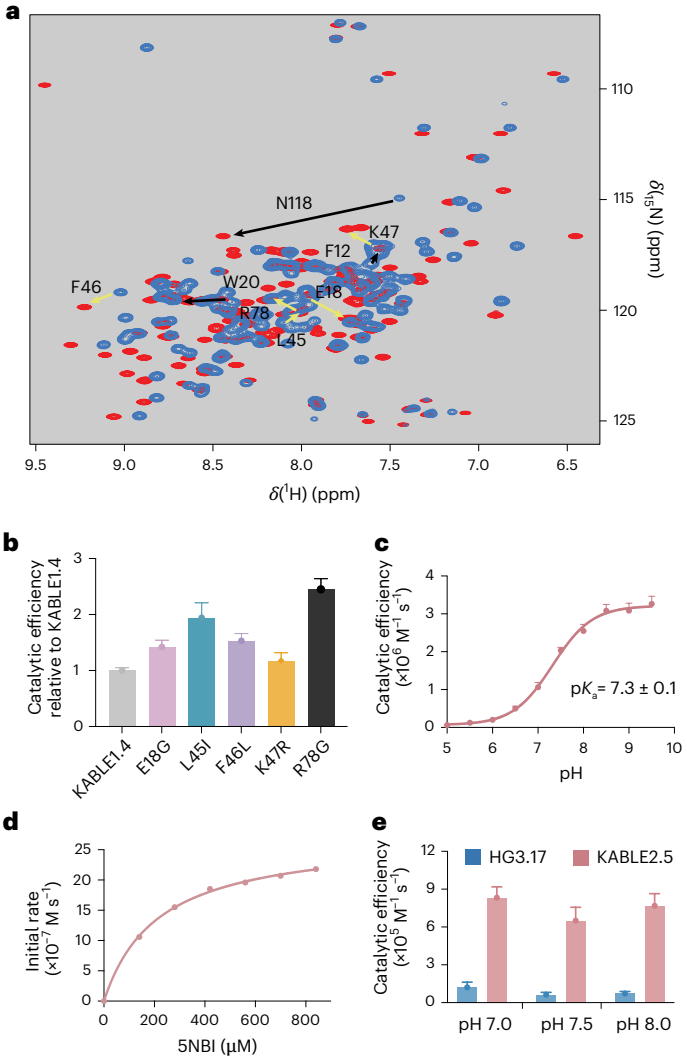


Fig. 4 | NMR-guided directed evolution of KABLE1.4. a, ^1H–^{15}N HSQC spectrum of KABLE1.4 in the absence (blue peaks) and presence (red peaks) of a transition state analogue (5 mol equiv. 6NBT). Black arrows represent the beneficial residues identified from a previous round of active site mutagenesis and yellow arrows represent residues with large chemical shift perturbations, where the second set of beneficial mutations were found. **b**, Catalytic efficiency of KABLE1.4 and its beneficial mutants. Kinetics were measured in pH 8.0 buffer (20 mM TRIS, pH 8.0, 100 mM NaCl, 1.5% acetonitrile) using a plate reader, following the increase in absorption of the product at 380 nm. Values are expressed as fold-change relative to KABLE1.4, which is set to 1. **c**, pH dependence of k_{cat}/K_M for KABLE2.5 measured by stopped-flow UV–vis spectroscopy. The k_{cat}/K_M values obtained at different pH were used to fit $(k_{cat}/K_M)_{max}$ and pK_a. **d**, Michaelis–Menten curve of KABLE2.5 measured by stopped-flow UV–vis spectroscopy in pH 8.0 buffer (50 mM Bis-Tris propane, 100 mM NaCl and 1.5% acetonitrile). **e**, Bar plot representation of the catalytic efficiency of KABLE2.5 and HG3.17. The activities for both proteins were measured using a plate reader under previously reported conditions to measure the activity of HG3.17 (ref. 45). pH 7.0 conditions: 50 mM sodium phosphate, 100 mM NaCl, 10% methanol and 27 °C; pH 7.5 conditions: 50 mM Bis-tris propane, 100 mM NaCl, 10% methanol and 27 °C; pH 8.0 conditions: 50 mM Bis-tris propane, 100 mM NaCl, 10% methanol and 27 °C. The error bars in **b**,**e** and **c**,**d** represent the standard errors of the mean from three and five independent measurements, respectively. Error bars in **d** are smaller than the symbol size and are therefore not visible.

Under the same conditions that had been optimized for HG3.17 activity (10% methanol), we found that the k_{cat}/K_M value for KABLE2.5 was 7- to 11-fold greater than HG3.17 between pH 7 and pH 8 (Fig. 4e, Supplementary Fig. 18 and Supplementary Table 7). KABLE2.5 was

https://doi.org/10.1038/s41557-026-02125-6

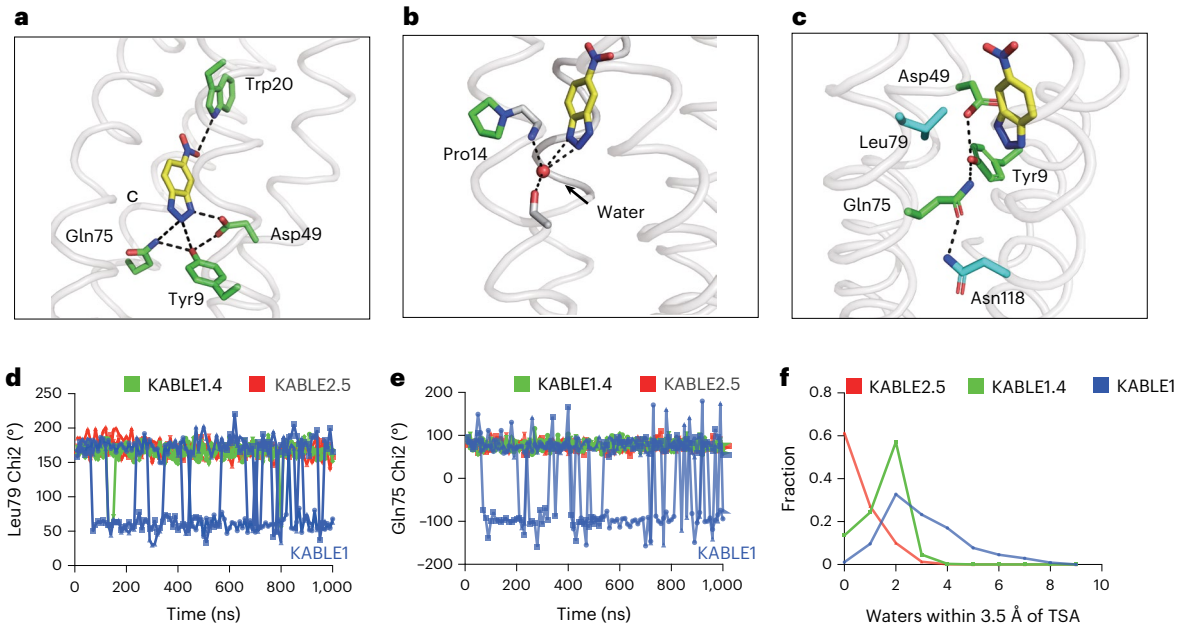


Fig. 5 | The origin of KABLE's efficient catalysis probed by X-ray crystallography and MD simulation. a, Structure of the KABLE1.2$_{cryst}$–6NBT complex (PDB 9N0J) showing the Asp triad engaging the triazole and a Trp indole NH engaging with the nitro group at the opposite end of the bound 6NBT. **b**, Structure of the KABLE1.2$_{cryst}$–6NBT complex (PDB 9N0J). The helical geometry is disrupted by Pro14, a backbone amide that binds a water molecule, which, in turn, forms two additional interactions with the triazole. **c**, MD snapshot showing the structure of the KABLE1.4–6NBT complex showing the Asp49-Tyr9-Gln75 triad (carbon atoms shown in green) and key residues, Leu79 and Asn118 (carbon atoms shown in blue; the carbon atoms of 6NBT are in yellow). **d**,**e**, Fluctuation of the Chi2 angle of Leu79 (**d**) and Gln75 (**e**) versus time during MD simulations. KABLE1 sampled two distinct conformers, only one of which is found in the more active KABLE1.4 and KABLE2.5 trajectories. Three independent simulations are shown for each system. **f**, The number of waters within 3.5 Å of the TSA. TSA, 6-nitrobenzotriazole.

also found to be an order of magnitude more active than the recently reported Kemp eliminase Des27.7 F113L (ref. 61), built using the starting natural protein scaffold from the same family of HG3.17, but with only two iterative rounds of computational design and experimental characterization, along with another serendipitous mutation (Fig. 3c).

Beneficial mutants immobilize and dehydrate the catalytic site

To understand the efficient catalysis of KABLE mutants, we determined structures by X-ray crystallography. Although the high water solubility of KABLE variants inhibited crystallization, substitutions of surface residues[12] enabled crystallization of the K14P and L20W mutants (designated KABLE1.2$_{cryst}$; Supplementary Fig. 19). The crystallographic structures at 1.6 and 2.2 Å resolution solved in the presence and absence of 6NBT, respectively, confirmed the overall structure and arrangement of the Asp triad (0.8 Å Cα RMSD relative to the designed KABLE1.2; Supplementary Table 8 and Supplementary Fig. 20): Asp49 formed a strong interaction (O–N distance of 2.5 Å) with the NH group of the transition state analogue, Gln75 formed a hydrogen bond with 6NBT's ring NH and Tyr9 simultaneously interacted with Gln75, Asp49 and 6NBT (Fig. 5a). The structure also provided a rationale for the beneficial K14P and L20W mutants discovered from saturation mutagenesis. The indole of L20W formed a weak hydrogen bond (3.3 Å) with the nitro group of 6NBT. Furthermore, the helix-breaking Pro14 locally disrupts the α-helix, enabling a backbone carbonyl to form a strong hydrogen bond with a water molecule, which interacts with the transition state analogue (Fig. 5b), similar to other Kemp eliminases[45].

We performed three independent microsecond molecular dynamics (MD) simulations of the 6NBT complexes of KABLE1, KABLE1.4 and KABLE2.5. Large changes in the dynamics were observed between KABLE1 and KABLE1.4, which has 170-fold greater activity (Extended Data Table 1). Three active site residues, Gln75, Leu79 and Asn118, had greater mobility in KABLE1 than in the more active KABLE1.4 (Fig. 5c–e and Supplementary Fig. 21): Leu79 interacts with the aromatic ring of 6NBT (Fig. 5d), Gln75 is a member of the catalytic triad (Fig. 5e) and Asn118 forms a hydrogen bond with Gln75 of the catalytic triad (Fig. 5c and Supplementary Fig. 21). Thus, the active conformation of the Asp triad is more rigid in KABLE1.4, suggesting that beneficial mutants quench non-productive dynamic fluctuations in KABLE1. The MD simulations also confirmed the role of the two other beneficial mutants seen in the crystal structure of KABLE1.2$_{cryst}$, including the water-mediated interaction near Pro14 seen in the crystal structure (Supplementary Fig. 22). Similar decreases in flexibility were observed in KABLE2.5 compared with KABLE1. Previous work has shown that dehydration of a carboxylate base is important for Kemp eliminase activity[62–64]. Indeed, the mean number of water molecules in contact with 6NBT decreased markedly from KABLE1 to KABLE2.5 (Fig. 5f). Together, these findings provide strong support for our proposed catalytic mechanism and rationalize the improvements in activity that were observed during the two rounds of mutagenesis.

Discussion

Proteins evolve on a landscape that is often portrayed pictorially by projecting the features of sequence space into two dimensions[2,65,66]. In this study, we instead used a similar landscape to represent both the protein sequence and chemical structure in designed proteins (Fig. 6a). We began with ABLE, which is specific for binding apixaban, progressed through an investigation of this same sequence's low-level chemical promiscuity and ultimately used computational design to reach two sequences, KABLE2.5 and FABLE, with distinct activities (Fig. 6a). Searching broadly and simultaneously through both chemical and sequence space requires the exploration of two vast molecular landscapes. Most previous experimental work has focused on exploring just one of these two variables: drug discovery involves an extensive search of chemical space to discover small molecules that bind to a target protein with a fixed sequence[67–69], while methods such as yeast display, phage display or directed evolution explore

https://doi.org/10.1038/s41557-026-02125-6

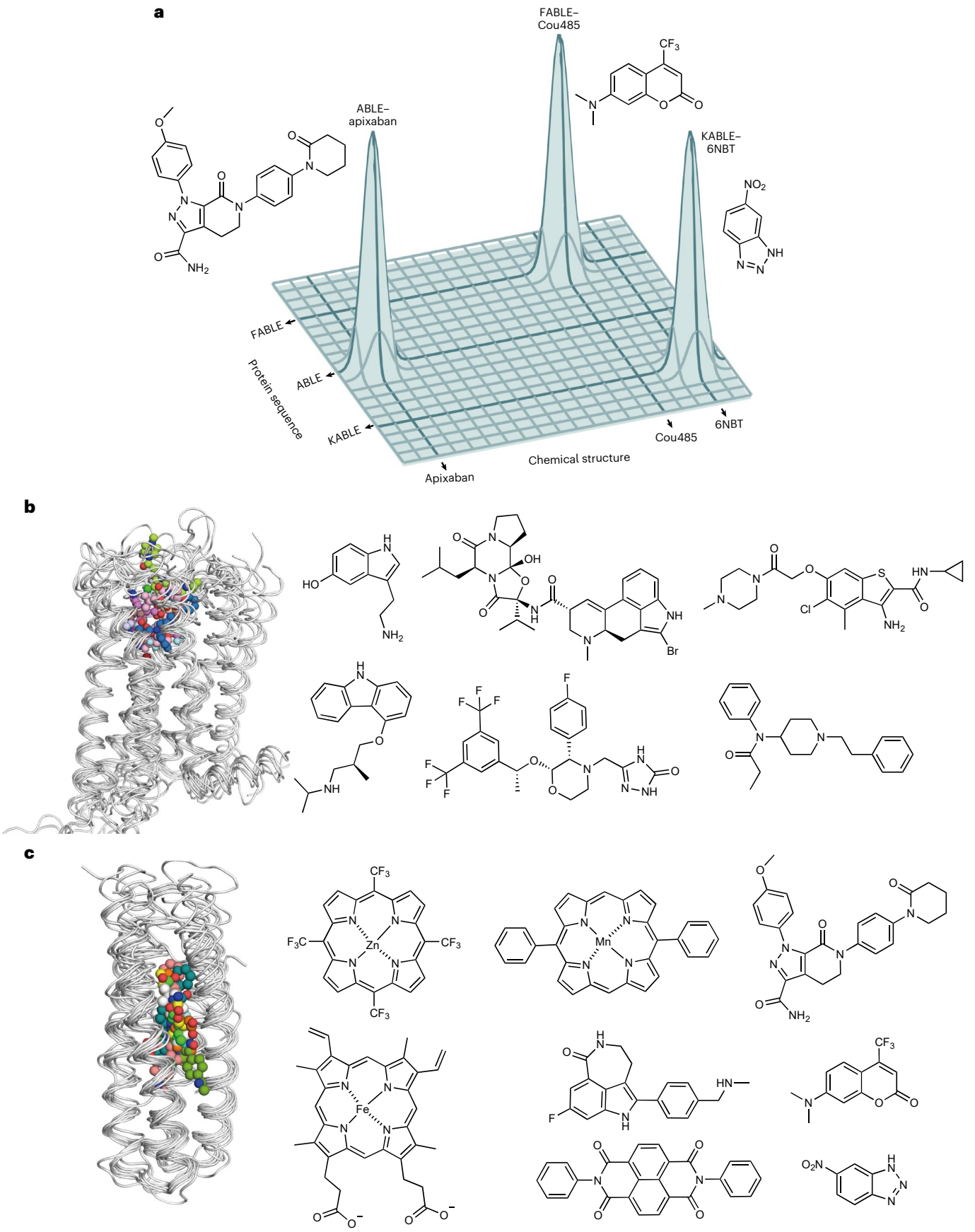


Fig. 6 | Exploration of sequence and chemical space of de novo protein–ligand interactions for the design of distinctive functions. a, A three-dimensional landscape showing protein sequence diversity and small-molecule chemical diversity on the horizontal axes, with the heights of peaks denoting fitness for binding small molecules or catalysing the Kemp elimination reaction. Crystallographic fragment screening revealed ABLE to be a generalist for binding 43 chemical fragments, which informed the design of the binding protein FABLE and the enzyme KABLE. **b**,**c**, Similar to GPCRs, the four-helical bundle scaffold has been used to bind different ligands: overlay of GPCR complexes and their diverse ligands (PDB 7WC4, 2RH1, 6ME9, 6OIK, 6VMS, 6HLO and 8EF5) (**b**) and de novo designed four-helical bundle complexes and their diverse ligands (PDB 5TGY/7JH6, 7JRQ, 6W70, 7AH0/8D9P, 8TN6, 9DWC, 9DM3 and 9N0J) (**c**).

sequence space to discover re-engineered protein sequences for a given small molecule[65,66]. De novo protein design provides an alternative approach to explore sequence space. Although de novo design of small-molecule-binding proteins is still considered a significant challenge[70], a number of proteins have been designed to bind drugs and cofactors[8,12,17,22,70–73]. However, it was previously unknown whether designed small-molecule-binding proteins are more or less specific than natural proteins. De novo proteins have great conformational stability[12] and their designed substrate-binding interactions are generally idealized. Together, these features might increase their specificity, making it difficult for them to act as generalists that can perform different functions.

Alternatively, they might have relatively low specificity because their binding sites have not been subject to the negative selective pressures that enforce specificity in naturally evolved proteins. Here, we found that the de novo protein ABLE had promiscuity similar to that of natural proteins and that its promiscuous low-level binding could be used as a starting point for designing specific binding and catalytic functions.

In drug discovery, fragment screening is used to discover chemotypes that bind to a target's active site. Such hits can guide the design of specialized inhibitors[67]. Here, we used fragment chemotype discovery to instead steer the design of proteins with distinct activities. Although we have used crystallographic fragment screening[20,35], there are many experimental and computational methods[67,68] that could be applied to protein design.

The facile emergence of Kemp eliminase activity in the ABLE scaffold is noteworthy. Kemp eliminases have previously been elaborated starting with the active sites of natural proteins, but the attainment of high activity was achieved after multiple rounds of computationally guided redesign[61] or directed evolution[45,46,74] (Fig. 3c). Despite these efforts, KABLE2.5 has at least a sevenfold greater value of k_{cat}/K_M than any previously reported base-catalysed Kemp eliminase. The differences in the ease of obtaining a targeted enzymatic activity from a de novo versus a natural protein might reflect the fact that natural enzymes have been selected to not only promote their cognate reactions but also avoid catalysing off-target reactivities that might be detrimental to an organism's survival. With a de novo scaffold, we begin with an open book without an evolutionary history predisposing towards one specific reactivity. It also frees one to match the scaffold with the reaction of interest[75–77]. The active site carboxylate of the KABLE1 series of proteins is located within a deep cavity of the protein, which enhances its dehydration and increases its basicity[62]. The enhancement of reactivity occurs at the expense of the stability of the protein as the burial of a carboxylate in a hydrophobic environment is thermodynamically unfavourable. The kinetically derived pK_a of Asp49 is 2.5 units higher than the unperturbed pK_a of Asp in water, corresponding to a free energy cost of approximately 3 kcal mol^{-1}.

Natural proteins are marginally stable and hence cannot easily withstand the multiple modifications required to generate an entirely different function. By contrast, the extreme stability of de novo proteins[3,78] enables such modifications. The high stability of KABLE2.5 (melting temperature >95 °C; Supplementary Fig. 22) also allowed the incorporation of multiple substitutions that were functionally beneficial but structurally destabilizing: K14P introduced a substitution that destabilizes α-helices by approximately 3 kcal mol^{-1} (ref. 79) and L118N replaced a hydrophobic interaction with a hydrogen-bonded interaction, which tends to be energetically unfavourable. Despite these changes, KABLE1.4 remained highly stable (Supplementary Fig. 23).

In pioneering work, Kuhlman and co-workers designed a Zn(II)-binding protein that catalyses ester hydrolysis[14,15] and a Diels–Alder reaction[16]. It is hard to directly compare this protein with KABLE2.5 because their catalytic activities rely on a metal ion cofactor. Nevertheless, it is striking that our protein remained much closer in structure to the original design[8,14–16], and the optimization of the FABLE and KABLE proteins was facilitated by the higher stability of the starting scaffold.

Nature has long used helical bundles as frameworks for enzymes that catalyse some of the most essential reactions for life, such as O_2 formation and utilization[32,80]. They are also the most widely used small-molecule-binding and -sensing devices. GPCRs bind small molecules of widely varying polarity, charge and hydrophobicity in a site that is bounded by helical bundles[28] (Fig. 6b). Although their sequences are highly diverse, their structures are remarkably well conserved (Supplementary Table 9)[81–83]. Similarly, de novo helical bundles represent a privileged scaffold for the synthetic recognition of a wide range of small molecules, which include haems[84–86], synthetic cofactors[38,87,88] and small-molecule drugs[12], in addition to the molecules studied here (Fig. 6c). Indeed, there is high similarity between the structures of ten de novo bundles designed in several different labs to bind a diverse array of small molecules (Supplementary Table 10). Thus, in de novo design as in Nature, form follows function.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-026-02125-6.

References

1. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
2. Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582 (2010).
3. Regan, L. & DeGrado, W. F. Characterization of a helical protein designed from first principles. *Science* **241**, 976–978 (1988).
4. Walsh, S. T. R., Cheng, H., Bryson, J. W., Roder, H. & DeGrado, W. F. Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc. Natl Acad. Sci. USA* **96**, 5486–5491 (1999).
5. Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
6. Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
7. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
8. Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Q. Rev. Biophys.* **53**, e3 (2020).
9. Kortemme, T. De novo protein design—from new structures to programmable functions. *Cell* **187**, 526–544 (2024).

10. Albanese, K. I., Barbe, S., Tagami, S., Woolfson, D. N. & Schiex, T. Computational protein design. *Nat. Rev. Methods Primers* **5**, 13 (2025).
11. Qing, R. et al. Protein design: from the aspect of water solubility and stability. *Chem. Rev.* **122**, 14085–14179 (2022).
12. Lu, L. et al. De novo design of drug-binding proteins with predictable binding energy and specificity. *Science* **384**, 106–112 (2024).
13. Yeh, A. H.-W. et al. De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
14. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940 (2012).
15. Studer, S. et al. Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).
16. Basler, S. et al. Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).
17. Stenner, R., Steventon, J. W., Seddon, A. & Anderson, J. L. R. A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. *Proc. Natl Acad. Sci. USA* **117**, 1419–1428 (2020).
18. Schnettler, J. D. et al. Selection of a promiscuous minimalist cAMP phosphodiesterase from a library of de novo designed proteins. *Nat. Chem.* **16**, 1200–1208 (2024).
19. Collins, P. M. et al. Achieving a good crystal system for crystallographic X-ray fragment screening. *Methods Enzymol.* **610**, 251–264 (2018).
20. Schuller, M. et al. Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).
21. Pearce, N. M. et al. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).
22. Polizzi, N. F. & DeGrado, W. F. A defined structural unit enables de novo design of small-molecule–binding proteins. *Science* **369**, 1227–1233 (2020).
23. Aharoni, A. et al. The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2005).
24. Bar-Even, A. et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).
25. Putman, S. J., Coulson, A. F., Farley, I. R., Riddleston, B. & Knowles, J. R. Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochem. J.* **129**, 301–310 (1972).
26. Roland, B. P. et al. Triosephosphate isomerase I170V alters catalytic site, enhances stability and induces pathology in a *Drosophila* model of TPI deficiency. *Biochim. Biophys. Acta Mol. Basis Dis.* **1852**, 61–69 (2015).
27. Rodríguez-Bolaños, M., Miranda-Astudillo, H., Pérez-Castañeda, E., González-Halphen, D. & Perez-Montfort, R. Native aggregation is a common feature among triosephosphate isomerases of different species. *Sci. Rep.* **10**, 1338 (2020).
28. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
29. Pillai, A. S. et al. Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).
30. Huang, J., Pan, X. & Yan, N. Structural biology and molecular pharmacology of voltage-gated ion channels. *Nat. Rev. Mol. Cell Biol.* **25**, 904–925 (2024).
31. Nishi, T. & Forgac, M. The vacuolar (H^+)-ATPases—nature's most versatile proton pumps. *Nat. Rev. Mol. Cell Biol.* **3**, 94–103 (2002).
32. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537**, 644–648 (2016).
33. Nozawa, K. et al. Pyrrolysyl-tRNA synthetase–tRNAPyl structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163–1167 (2009).
34. Williams, P. A. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**, 464–468 (2003).
35. Gahbauer, S. et al. Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2. *Proc. Natl Acad. Sci. USA* **120**, e2212931120 (2023).
36. Aschenbrenner, J. C. et al. Identifying novel chemical matter against the Chikungunya virus nsP3 macrodomain through crystallographic fragment screening. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.23.609196 (2024).
37. Nad, S. & Pal, H. Unusual photophysical properties of coumarin-151. *J. Phys. Chem. A* **105**, 1097–1106 (2001).
38. Polizzi, N. F. et al. De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
39. Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
40. Wu, R. et al. High-resolution de novo structure prediction from primary sequence. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.21.500999 (2022).
41. Cao, D. et al. Coumarin-based small-molecule fluorescent chemosensors. *Chem. Rev.* **119**, 10403–10519 (2019).
42. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA* **96**, 9997–10002 (1999).
43. Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **13**, 105–121 (2014).
44. Röthlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
45. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418–421 (2013).
46. Patsch, D. et al. Enriching productive mutational paths accelerates enzyme evolution. *Nat. Chem. Biol.* **20**, 1662–1669 (2024).
47. Privett, H. K. et al. Iterative approach to computational enzyme design. *Proc. Natl Acad. Sci. USA* **109**, 3790–3795 (2012).
48. Korendovych, I. V. et al. Design of a switchable eliminase. *Proc. Natl Acad. Sci. USA* **108**, 6823–6827 (2011).
49. Bhattacharya, S. et al. NMR-guided directed evolution. *Nature* **610**, 389–393 (2022).
50. Broom, A. et al. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).
51. Risso, V. A. et al. De novo active sites for resurrected Precambrian enzymes. *Nat. Commun.* **8**, 16113 (2017).
52. Lamba, V. et al. Kemp eliminase activity of ketosteroid isomerase. *Biochemistry* **56**, 582–591 (2017).
53. Bunzel, H. A. et al. Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat. Chem.* **13**, 1017–1022 (2021).
54. Schwans, J. P., Kraut, D. A. & Herschlag, D. Determining the catalytic role of remote substrate binding interactions in ketosteroid isomerase. *Proc. Natl Acad. Sci. USA* **106**, 14271–14275 (2009).
55. Gandour, R. D. On the importance of orientation in general base catalysis by carboxylate. *Bioorg. Chem.* **10**, 169–176 (1981).
56. Dauparas, J. et al. Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **22**, 717–723 (2025).
57. Khersonsky, O. et al. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **396**, 1025–1042 (2010).

58. Jing, X. et al. Single-sequence protein structure prediction by integrating protein language models. *Proc. Natl Acad. Sci. USA* **121**, e2308788121 (2024).
59. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
60. Merlicek, L. P. et al. AI.zymes: a modular platform for evolutionary enzyme design. *Angew. Chem. Int. Ed.* **64**, e202507031 (2025).
61. Listov, D. et al. Complete computational design of high-efficiency Kemp elimination enzymes. *Nature* **643**, 1421–1427 (2025).
62. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294–2301 (1973).
63. Frushicheva, M. P., Cao, J., Chu, Z. T. & Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proc. Natl Acad. Sci. USA* **107**, 16869–16874 (2010).
64. Wang, P., Zhang, J., Zhang, S., Lu, D. & Zhu, Y. Using high-throughput molecular dynamics simulation to enhance the computational design of Kemp elimination enzymes. *J. Chem. Inf. Model.* **63**, 1323–1337 (2023).
65. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
66. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
67. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).
68. Chen, Y. & Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **5**, 358–364 (2009).
69. Bedard, P. L., Hyman, D. M., Davids, M. S. & Siu, L. L. Small molecules, big impact: 20 years of targeted therapy in oncology. *Lancet* **395**, 1078–1088 (2020).
70. An, L. et al. Binding and sensing diverse small molecules using shape-complementary pseudocycles. *Science* **385**, 276–282 (2024).
71. Yang, W. & Lai, L. Computational design of ligand-binding proteins. *Curr. Opin. Struct. Biol.* **45**, 67–73 (2017).
72. Tinberg, C. E. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
73. Dou, J. et al. De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
74. Gutierrez-Rus, L. I. et al. Enzyme enhancement through computational stability design targeting NMR-determined catalytic hotspots. *J. Am. Chem. Soc.* **147**, 14978–14996 (2025).
75. Tantillo, D. J., Chen, J. & Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **2**, 743–750 (1998).
76. Frushicheva, M. P. et al. Computer aided enzyme design and catalytic concepts. *Curr. Opin. Chem. Biol.* **21**, 56–62 (2014).
77. Acosta-Silva, C., Bertran, J., Branchadell, V. & Oliva, A. Kemp elimination reaction catalyzed by electric fields. *ChemPhysChem* **21**, 295–306 (2020).
78. Huang, P.-S. et al. High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
79. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**, 422–427 (1998).
80. Bhowmick, A. et al. Structural evidence for intermediates during O_2 formation in photosystem II. *Nature* **617**, 629–636 (2023).
81. Cao, D. et al. Structure-based discovery of nonhallucinogenic psychedelic analogs. *Science* **375**, 403–411 (2022).
82. Yin, J. et al. Structure of a D2 dopamine receptor–G-protein complex in a lipid membrane. *Nature* **584**, 125–129 (2020).
83. Cherezov, V. et al. High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
84. Choma, C. T. et al. Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116**, 856–865 (1994).
85. Hutchins, G. H. et al. An expandable, modular de novo protein platform for precision redox engineering. *Proc. Natl Acad. Sci. USA* **120**, e2306046120 (2023).
86. Ennist, N. M., Stayrook, S. E., Dutton, P. L. & Moser, C. C. Rational design of photosynthetic reaction center protein maquettes. *Front. Mol. Biosci.* **9**, 997295 (2022).
87. Mann, S. I., Nayak, A., Gassner, G. T., Therien, M. J. & DeGrado, W. F. De novo design, solution characterization, and crystallographic structure of an abiological Mn–porphyrin-binding protein capable of stabilizing a Mn(V) species. *J. Am. Chem. Soc.* **143**, 252–259 (2021).
88. Mann, S. I. et al. De novo design of proteins that bind naphthalenediimides, powerful photooxidants with tunable photophysical properties. *J. Am. Chem. Soc.* **147**, 7849–7858 (2025).
89. Khersonsky, O. et al. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl Acad. Sci. USA* **109**, 10358–10363 (2012).

Methods

ABLE crystallization and fragment screening

Lyophilized ABLE was dissolved in water to a concentration of 18 mg ml^{-1}. Crystals were grown by sitting-drop vapour diffusion in SwissCI three-well plates (HR3-125, Hampton) with 200 nl protein and 200 nl reservoir (220 mM sodium malonate, pH 5, 20% PEG3350 and 3% dimethyl sulfoxide (DMSO)). Needle-shaped crystals grew overnight at 20 °C. We screened the 320 compound Enamine Essential Fragment Library against ABLE. Fragment solutions (40 nl of 100 mM stock solutions prepared in DMSO) were added to crystal drops using an acoustic liquid handler (Echo 650, Beckman Coulter)[90]. Crystals were incubated with fragments for 24 h before being looped and vitrified in liquid nitrogen. X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source and beamlines 12-1 and 12-2 of the Stanford Synchrotron Radiation Lightsource. The data collection strategies are summarized in Supplementary Information.

Diffraction images were indexed, integrated and scaled with XDS(version 30 June 2023) and merged with Aimless (version 0.8.2)[91,92]. In total, we mounted 447 crystals soaked with fragments and were able to obtain high-quality datasets from 321 crystals (resolution limit <2.2 Å based on $CC_{1/2} < 0.3$, R_{free} of initial model <35%). $CC_{1/2}$ is the Pearson correlation coefficient between the average intensities of corresponding unique reflections randomly split into two datasets. Replicate soaks were performed for some fragments: the total number of unique fragments with high-quality datasets was 242 (Supplementary Information). We initially refined a model of ABLE against a dataset collected from a crystal soaked only in 10% DMSO. Phases were obtained by molecular replacement using Phaser[93] and chain A of the previously published ABLE X-ray crystal structure as the search model (PDB 6W6X)[22]. The initial model was improved by iterative cycles of refinement with phenix.refine (version 1.21.1-5286)[94] and manual model building with Coot[95]. Water molecules were placed automatically into peaks in the $F_o - F_c$ difference map >3.5σ in the early stages of refinement using phenix.refine. In the later stages of refinement, water molecules were added or deleted manually, and hydrogen atoms were refined with parameters constrained by those of the non-hydrogen atom (the riding hydrogen model). Data collection and refinement statistics are reported in Supplementary Information. Coordinates and structure factor amplitudes have been deposited in the PDB with accession code 9DW2. Fragment-soaked datasets were refined using a pipeline based on Dimple as described previously using the DMSO-soaked ABLE model (PDB 9DW2)[96,97]. The pipeline consisted of initial rigid body refinement, followed by two cycles of restrained refinement in Refmac (version 5.8.0258): the first with harmonic distance restraints (jelly body restraints, four cycles) and the second with restrained refinement (eight cycles).

Fragment binding was detected using the Pan Dataset Density Analysis (PanDDA) algorithm[21] packaged in CCP4 (version 7.0)[98]. The background electron density map was calculated from 30 datasets (Supplementary Information): six datasets were from crystals soaked in 10% DMSO and the remaining came from fragment-soaked crystals in which no fragment was detected. We collected multiple datasets for selected fragments. If fragments were detected in multiple datasets, we modelled the fragment with the highest occupancy based on the PanDDA event map and background density correction (1-BDC) value, where BDC is the fraction of background map subtracted from the fragment-bound map. Fragments were modelled into PanDDA event maps (Supplementary Information) using Coot (version 0.8.9.2)[95] and changes in protein conformation and solvent near the fragments were modelled. Fragment restraint files were generated with phenix.elbow[99] from three-dimensional coordinates generated by LigPrep (version 2022-1, Schrödinger) or Grade 2 (Global Phasing). Structure refinement was performed with phenix.refine (version 1.20.1-4487) starting from the ABLE DMSO reference structure (PDB 9DW2). Alternative conformations were modelled for residues when the RMSD exceeded 0.15 Å from the apo structure. The apo conformation was assigned alternative location (altloc) A and the fragment-bound conformation was assigned altloc B. Two fragments bound with alternative conformations that overlapped with the Tyr46 side chain (PDBs 7HJK and 7HJX). For these fragments, residues 44–48 were modelled with a second conformation (altloc C). The multistate models were initially refined with five phenix.refine macrocycles without hydrogen atoms, followed by ten macrocycles with riding hydrogen atoms. Occupancy of the fragment and apo states were refined at either 2 × (1-BDC) (refs. 21,100) or at fragment occupancy values from 10–90% at 10% increments. Fragment occupancy was determined by inspection of $mF_o - DF_c$ difference map peaks after refinement. The occupancy at 2 × (1-BDC) was appropriate for 41/43 fragments (Supplementary Information). Coordinates and structure ×factor amplitudes have been deposited in the PDB with the accession codes 7HIY, 7HIZ, 7HJ0, 7HJ1, 7HJ2, 7HJ3, 7HJ4, 7HJ5, 7HJ6, 7HJ7, 7HJ8, 7HJ9, 7HJA, 7HJB, 7HJC, 7HJD, 7HJE, 7HJF, 7HJG, 7HJH, 7HJI, 7HJJ, 7HJK, 7HJL, 7HJM, 7HJN, 7HJO, 7HJP, 7HJQ, 7HJR, 7HJS, 7HJT, 7HJU, 7HJV, 7HJW, 7HJX, 7HJY, 7HJZ, 7HK0, 7HK1, 7HK2, 7HK3 and 7HK4. Structure factor intensities (unmerged, merged and merged/scaled), PanDDA input and output files (including Z-map and event maps in CCP4 format), and refined models (including the fragment-bound state extracted from multistate models) are available via Zenodo at https://doi.org/10.5281/zenodo.13913848.

Computational design of FABLE

With the (7-hydroxycoumarin)-ABLE structure (PDB 7HIY) available from fragment screening, we superimposed Cou485 on the 7-hydroxycoumarin fragment based on shared heavy atoms to obtain the initial pose, which includes Tyr46 packing with the 2*H*-1-benzopyran ring of coumarin and the hydrogen bond between His49 and the ketone of coumarin. This initial pose was used for Rosetta flexible backbone sequence design using a previously reported script[38] (Supplementary Note and https://doi.org/10.5281/zenodo.17935960). During the sequence design, the side chain conformations of His49 and Tyr46 were fixed and the geometry constraint for the His49 interaction with the ketone of Cou485 was applied. Of the total 126 residues, 43 residues in the vicinity of apixaban and 7-hydroxycoumarin binding site were chosen to be designed to reshape the binding pocket towards Cou485 while keeping the portion of the packing core of ABLE that is not involved in binding ligands unchanged. We selected the 100 best-scoring designs (lowest Rosetta energy score, using the Rosetta Ref2015 energy function) from 1,000 design outputs. We next predicted the structure using OmegaFold[40] and ESMFold[59] to further filter out the designs whose predicted structures do not agree with the designed structures. Sequences with an RMSD difference between the design model and prediction less than 1.0 Å as well as a predicted local distance difference test score over 90 were picked for detailed structural evaluation. The final sequences were selected using two additional criteria: (1) no more than two polar atoms at side chains of TSA-interacting residues without hydrogen bond and (2) a preorganized binding pocket from the prediction, especially the side-chain conformations of Tyr46 and His49, that is quantified by sub-ångström RMSDs for Cα, Cβ and Cγ. between the designed model and AlphaFold 3 prediction. The AlphaFold 3 server (https://alphafoldserver.com) was later used to predict the structures of selected designs (Supplementary Fig. 6), although the program was not available at the time that the sequences were designed[101].

Computational design of KABLE (Kemp eliminase)

Our goal was to use the information from the fragment complexes to inform the design of the active site. As the substrate was a bicyclic aromatic compound, it was logical to position it in the aromatic boxes of the A site, hosting eight bicyclic fragments. This was followed by searching for a Glu or Asp side chain that was well positioned to abstract a proton from the substrate to catalyse the elimination reaction (Strategy 1). A second strategy centred instead on using the information from

fragment complexes to first identify a catalytically favourable position for a Glu/Asp transition state complex, followed by the redesign of ABLE's binding site to stabilize the bicyclic ring, primarily through hydrophobic aromatic interactions and van der Waals interactions (https://doi.org/10.5281/zenodo.17935960).

In each case, we considered theoretical suggestions that emphasize the location of a binding site relative to the bulk solvent. Specifically, it is important to consider placement of the base in a preorganized, solvent-inaccessible location well below the surface of the protein. This principle has been supported by a body of theoretical and computational studies, as well as through careful examination of evolved Kemp eliminases[45,47,48,63]. We satisfied Perutz's principle by choosing designs in which the carboxylate base and substrate were fully inaccessible to a water-sized probe, the surrounding side chains should be primarily hydrophobic[47] and the geometry of the carboxylate should be close to the saddle point for the computed transition state of the reaction.

Strategy 1 of using aromatic box. Inspired by the aromatic boxes of conformer A, we used a similar procedure for the design of a Kemp eliminase. For the aromatic box in conformer A, we used PDB structure 7HJQ as the starting structure. The plane of the ring of 6NBT was oriented as in PDB structure 7HK4. The next step was to identify a position for the incorporation of the active site Asp or Glu to satisfy the following criteria: (1) when in a low-energy rotamer, its carboxylate should form a hydrogen bond with the N3 atom of 6NBT in a geometry derived from previous quantum calculations[44,102,103]. (2) The benzotriazole ring of the substrate should be fully buried (to increase the basicity of the carboxylate). Thus, the Asp/Glu carboxylate should be inaccessible to solvent. We also filtered out any examples in which the side chain of the acidic residue points towards the solvent because the burial of the carboxylate would need to compete with solvation. (3) The introduced side chain should not clash with the backbone or side chains of the residues comprising the aromatic box. This left three positions to explore, namely, 108, 112 and 115. After searching through the low-energy side-chain rotamers of Asp and Glu at these locations, Glu at position 108 was found to uniquely fulfill these criteria. The resulting complex was used as an input structure for sequence design by either (1) an iterative round of LigandMPNN[56] and Rosetta FastRelax[39] or (2) a classical flexible backbone Rosetta sequence design:

1. For iterative rounds of LigandMPNN and Rosetta FastRelax, residue 46 was fixed to either Tyr or Trp and residue 86 was fixed to Phe to maintain the aromatic packing interactions discovered in fragment screening. In the Rosetta FastRelax step, geometric constraints from the reported crystallographic conformation sampling were applied (Supplementary Note, constrain for active site 1 of Kemp eliminase)[104].
2. The second route using the flexible backbone Rosetta sequence design proceeded with the same constraint applied (Supplementary Note, constrain for active site 1 of Kemp eliminase) following the reported protocol from our lab[38].

Structure prediction with RaptorX was carried out for all designed sequences[58,105] to ensure that the designed structures would be preorganized to adopt the same active site geometry of the design model. Of 2,000 sequences evaluated according to process (1), we chose designs 1, 4 and 5. Of 2,000 sequences evaluated according to process (2), we chose designs 2 and 3. The selection criteria were sub-ångström backbone RMSD between the designed complex and the predicted unliganded structures and also agreement between the design and prediction of the side-chain conformation of the active site (residues 108, 46, 79 and 86).

Strategy 2 of using hotspot residue 49. In strategy 2, we focused on identifying a site that is well positioned to accommodate a Glu/Asp as the general base. His49 was selected as a potentially ideal position because it is situated at a central location near the bottom of the binding site where it donates a hydrogen bond to a carbonyl of apixaban in ABLE. His49 also forms hydrogen bonds with 35% of the fragments binding in the pocket of ABLE and it forms imidazole–carboxylate salt bridges in nine of the complexes, as well as with acetate in PDB structure 6W6X (Supplementary Fig. 9). We reasoned that it should be possible to reverse the polarity of the interaction by replacing the His with an Asp, which is nearly isosteric with histidine, as this would place the side chain in a good environment to form a strong hydrogen bond. 6NBT was docked at the Asp49 of ABLE–fragment crystal structures (PDB 7HJV) in accord with previously reported geometric constraints from density functional theory calculations, considering the more basic (*syn*) lone pairs of electrons on either OD1 or OD2 for proton abstraction[44].

All three resulting poses (Supplementary Fig. 9) were used as input structures for sequence design with three iterative rounds of LigandMPNN and Rosetta FastRelax, resulting in 1,000 designs per input pose. In the Rosetta FastRelax step, geometric constraints from the reported crystallographic conformation sampling were applied (Supplementary Note, constrain for versatile residue 49 of Kemp eliminase)[104]. The structures of all designed sequences were predicted with RaptorX[58,105]. Five designs were picked according to the goodness of fit between the predicted model of the uncomplexed structure and the design model for the complex as described in the section 'Design of an efficient Kemp eliminase'.

MD simulations

General MD procedure. Proteins and ligands were parameterized using Gaussian09 and AmberTools's Antechamber program in Amber22[106,107]. The partial charges of ligands were determined by first optimizing their geometries at the B3LYP/6-31 G* level of theory[107] and then calculating their electrostatic potentials using the Merz–Singh–Kollman method in Gaussian09[107]. Charge fitting was then performed using Antechamber's RESP program within AmberTools[108,109]. All other small-molecule parameters were assigned by Antechamber based on the GAFF2 database[110,111]. The starting structures for the MD simulations are described in the sections 'MD for ABLE' and 'MD for KABLEs' The simulation box was built by solvating protein with 'optimal' point charge model of water molecules[112] in a box with 8 Å padding around the protein, and sodium and chloride ions were added to reach a charge-neutral NaCl concentration of 137 mM to match the experimental conditions. Simulations began with 1,000 restrained steepest-descent minimization steps before switching to a maximum of 6,000 steps in conjugate gradient steps. The system was then heated to 278 K over 50 ps in the NVT (number of particles, volume and temperature) ensemble with the temperature controlled with a Langevin thermostat and a time step of 1 fs. The simulation was then switched to the NPT (number of particles, pressure and temperature) ensemble and the pressure was maintained at 1 atm using the Monte Carlo barostat[113]. Throughout the equilibration steps, protein and ligand heavy atoms were initially restrained with harmonic potentials at 10 kcal mol^{-1} $Å^{-2}$ and then ramped down to 0 kcal mol^{-1} $Å^{-2}$ over nine equilibration steps totalling 1 ns. Each simulation was then carried out for an unrestrained production run under periodic boundary conditions with time steps of 2 fs. The SHAKE algorithm[114,115] was used to restrain hydrogen atoms, short-range non-bonded electrostatic and Lennard-Jones interactions were cut off at 10 Å, and the particle mesh Ewald method was used for long-range electrostatics.

MD for ABLE. The starting structures for the simulations were derived from the crystal structures of ABLE complexed with apixaban (PDB 6W70, chain B) and apo ABLE (PDB 6W6X, chain A)[22]. Both systems were prepared using the epsilon nitrogen-protonated state of H49. The simulation was carried out at 278 K with two independent runs,

each for 500 ns. The final trajectory was extracted with an interval of 5 ns, which resulted in 100 states for each run of 500 ns. The Chi1 and Chi2 angles of Tyr46 were extracted using an in-house script (Supplementary Note, extract_chi1_chi2.py) and manual inspection was also conducted to verify the correct calculation of the Chi2 angle due to the symmetry of the tyrosine side chain. The Chi1 and Chi2 angles of His49 were extracted using the same script. Contours were generated from the Chi1 and Chi2 angles observed in the simulation using an in-house script ('generate_contour_from_tyr_chi1and2.py' in Supplementary Note and https://doi.org/10.5281/zenodo.17935960), which uses the Gaussian kernel density estimate methods from the scipy.stats module of the SciPy Python packages.

MD for KABLEs. The starting structure for the KABLE1 simulations was derived from the design structures of KABLE1 complexed with 6NBT. For the simulation of KABLE1.4, its four mutations were incorporated using Rosetta FastRelax and the relaxed structures were used for the MD simulations. The simulations were carried out at 300 K for three independent runs using the Berendsen barostat, with each run lasting 1,000 ns. The final trajectory was extracted with an interval of 1 ns, which resulted in 1,000 states for each run of 1,000 ns. The Chi1 and Chi2 angles of Gln75 and Leu79 were extracted using an in-house script (Supplementary Note, extract_chi1_chi2.py). The number of water molecules within 3.5 Å of 6NBT was calculated using another script (Supplementary Note, count_water.py).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The crystal structures of ABLE–fragment complexes have been deposited in the PDB with accession codes 7HIY, 7HIZ, 7HJ0, 7HJ1, 7HJ2, 7HJ3, 7HJ4, 7HJ5, 7HJ6, 7HJ7, 7HJ8, 7HJ9, 7HJA, 7HJB, 7HJC, 7HJD, 7HJE, 7HJF, 7HJG, 7HJH, 7HJI, 7HJJ, 7HJK, 7HJL, 7HJM, 7HJN, 7HJO, 7HJP, 7HJQ, 7HJR, 7HJS, 7HJT, 7HJU, 7HJV, 7HJW, 7HJX, 7HJY, 7HJZ, 7HK0, 7HK1, 7HK2, 7HK3 and 7HK4. A summary of data collection, modelling and refinement of ABLE–fragment complexes is provided in Supplementary Information. Structure factor intensities (unmerged, merged and merged/scaled), PanDDA input and output files (including Z-map and event maps in CCP4 format), and refined models (including the fragment-bound state extracted from multistate models) are available via Zenodo at https://doi.org/10.5281/zenodo.13913848 (ref. 116). The crystal structures of FABLE and KABLE have been deposited in the PDB with accession codes 9DWA, 9DWB, 9DWC, 9N0I and 9N0J. X-ray data collection and refinement statistics for the FABLE and KABLE structures are provided in Supplementary Information. All other data not included in the article and Supplementary Information are available from the corresponding authors upon request. Source data are provided with this paper.

Code availability

Customized Python scripts used for protein design and modelling are provided in Supplementary Information. These Python scripts are also publicly available via Zenodo at https://doi.org/10.5281/zenodo.17935960 (ref. 117).

References

90. Collins, P. M. et al. Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr. D* **73**, 246–255 (2017).
91. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
92. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
93. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
94. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
95. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
96. Keegan, R., Wojdyr, M., Winter, G. & Ashton, A. DIMPLE: a difference map pipeline for the rapid screening of crystals on the beamline. *Acta Crystallogr. A* **71**, s18 (2015).
97. Correy, G. J. et al. Exploration of structure–activity relationships for the SARS-CoV-2 macrodomain fromshape-based fragment linking and active learning. *Sci. Adv.* **11**, eads7187 (2025).
98. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
99. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
100. Pearce, N. M., Krojer, T. & von Delft, F. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crystallogr. D* **73**, 256–266 (2017).
101. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
102. Alexandrova, A. N., Röthlisberger, D., Baker, D. & Jorgensen, W. L. Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J. Am. Chem. Soc.* **130**, 15907–15915 (2008).
103. Świderek, K., Tuñón, I., Moliner, V. & Bertran, J. Protein flexibility and preorganization in the design of enzymes. The Kemp elimination catalyzed by HG3.17. *ACS Catal.* **5**, 2587–2595 (2015).
104. Rakotoharisoa, R. V. et al. Design of efficient artificial enzymes using crystallographically enhanced conformational sampling. *J. Am. Chem. Soc.* **146**, 10001–10013 (2024).
105. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* **44**, W430–W435 (2016).
106. Case, D. A. et al. AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).
107. Gaussian 09, revision A.02. *ScienceOpen* https://www.scienceopen.com/document?vid=6be7271f-f651-464b-aee6-ef20b0743b6b (2009).
108. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
109. Besler, B. H., Merz, K. M. Jr & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).
110. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).
111. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
112. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models: a different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
113. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
114. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
115. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

116. Correy, G. J. & Fraser. J. S. PanDDA analysis of fragment screen against the computationally designed ABLE protein. *Zenodo* https://doi.org/10.5281/zenodo.13913848 (2025).
117. Chen, Y. Emergence of specific binding and catalysis from a designed generalist binding protein. *Zenodo* https://doi.org/10.5281/zenodo.17935960 (2025).
118. Thorn, S. N., Daniels, R. G., Auditor, M. T. & Hilvert, D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **373**, 228–230 (1995).
119. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* **383**, 60–63 (1996).
120. Khersonsky, O. et al. Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412 (2011).
121. Merski, M. & Shoichet, B. K. Engineering a model protein cavity to catalyze the Kemp elimination. *Proc. Natl Acad. Sci. USA* **109**, 16179–16183 (2012).
122. Gutierrez-Rus, L. I., Alcalde, M., Risso, V. A. & Sanchez-Ruiz, J. M. Efficient base-catalyzed Kemp elimination in an engineered ancestral enzyme. *Int. J. Mol. Sci.* **23**, 8934 (2022).
123. Vaissier, V., Sharma, S. C., Schaettle, K., Zhang, T. & Head-Gordon, T. Computational optimization of electric fields for improving catalysis of a designed Kemp eliminase. *ACS Catal.* **8**, 219–227 (2018).

Acknowledgements

We thank a reviewer for suggesting the direct comparison of HG3.17 with KABLE2.5. We are grateful for helpful discussions with S. Schneider and H. Jo. We thank members of the DeGrado and Fraser labs for support. The synchrotron X-ray diffraction data used to determine the crystal structures reported in this Article were collected at beamline 8.3.1 of the Advanced Light Source (ALS) and beamlines 12-1 and 12-2 of the Stanford Synchrotron Radiation Lightsource (SSRL). We acknowledge the use of the Wynton high-performance computer cluster at UCSF. We are grateful to the National Science Foundation (CHE-2108660 and MCB- 2306190, to W.F.D.) and the National Institutes of Health (R35GM122603, to W.F.D.) for funding. J.S.F. was supported by a Sanghvi-Agarwal Innovation Award and the National Institutes of Health (NIH GM145238). W.F.D. thanks the W. M. Keck Foundation for its support. Use of the SSRL, SLAC National Accelerator Laboratory, was supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (contract no. DE-AC02-76SF00515). The SSRL Structural Molecular Biology Program is supported by the US DOE, Office of Biological and Environmental Research and the National Institutes of Health, National Institute of General Medical Sciences (grant no. P30GM133894). The ALS, a US DOE Office of Science User Facility (contract no. DE-AC02-05CH11231), is supported in part by the ALS-ENABLE program funded by the NIH, National Institute of General Medical Sciences (grant no. P30GM124169). N.F.P. acknowledges funding from the NIH (R00GM135519). J.E.G. acknowledges funding from the NIH (R01GM141299). I.V.K. is thankful for funding support from the National Institute of Health (NIH R35GM119634) and Welch Foundation (AA-2198-20240404). S.B. is the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2522-24). S.B. is a recipient of a postdoctoral independent research grant (7032759) from the Program for Breakthrough Biomedical Research, which the Sandler Foundation partially funds. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Author contributions

Y.C., S.B., N.F.P., J.S.F. and W.F.D. formulated the project. L.B., G.J.C. and J.T.B. performed the fragment screening. Y.C. performed the computational design and carried out the experimental characterization of FABLEs. Y.C. conducted the computational design of KABLEs. S.B. and Y.C. carried out the experimental optimization and characterization of KABLE. S.B. conducted the directed evolution of KABLE. A.N.V. collected the NMR spectra. L.B., G.J.C. and Y.C. obtained the crystal structure data for the FABLEs and KABLEs. Y.C. and S.K.T. performed the MD simulations. K.H., L.L. and I.B. assisted with the computational design. Y.C., S.B., G.J.C., K.H., J.E.G., I.V.K., N.F.P., J.S.F. and W.F.D. wrote the paper with input from all authors.

Competing interests

J.S.F. has an equity in and is a compensated consultant for Profluent Bio. The other authors declare no competing interests.

Additional information

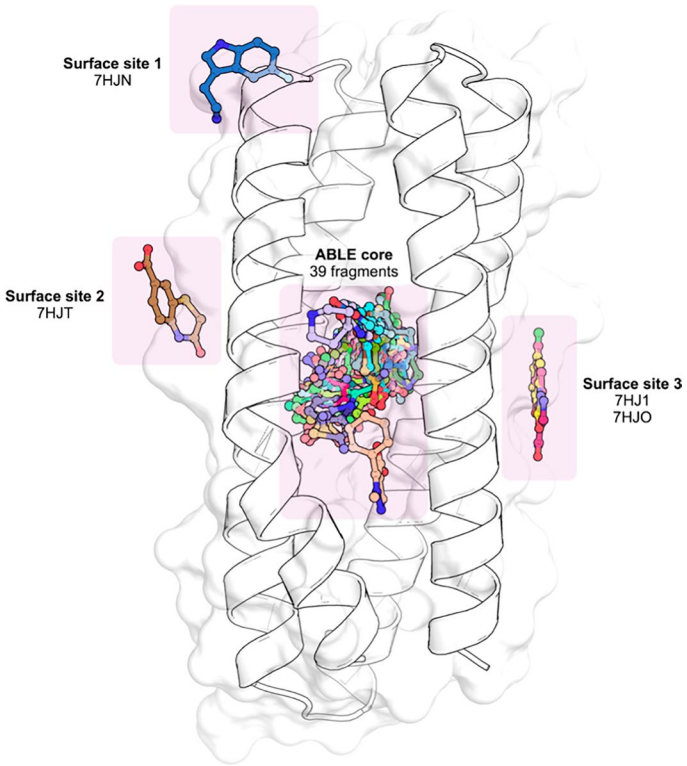
Extended data is available for this paper at https://doi.org/10.1038/s41557-026-02125-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41557-026-02125-6.

Correspondence and requests for materials should be addressed to James S. Fraser or William F. DeGrado.

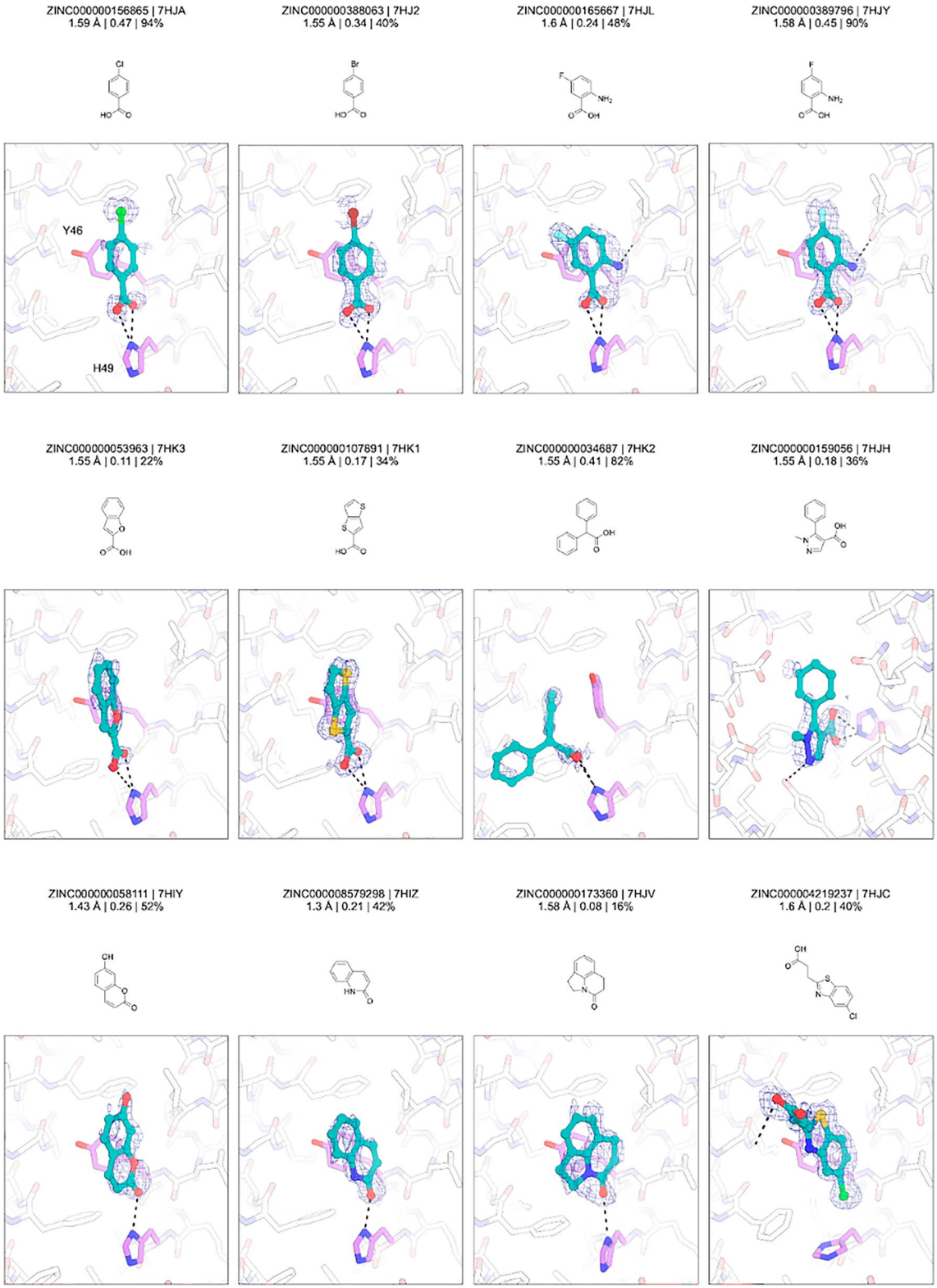
Peer review information *Nature Chemistry* thanks Shuguang Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41557-026-02125-6



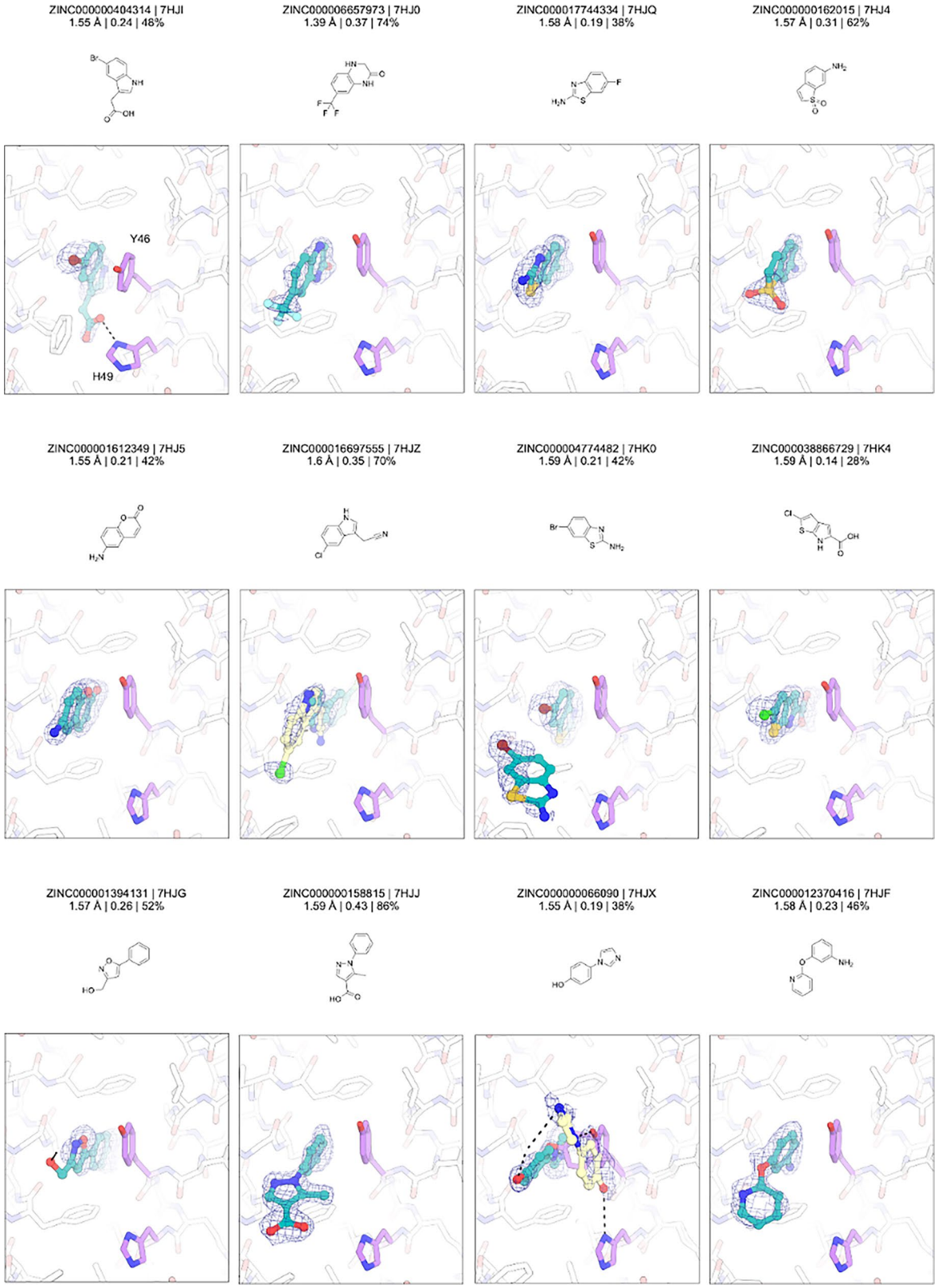
Extended Data Fig. 1 | X-ray crystal structure of ABLE showing binding sites for 43 fragments. The structure of apo ABLE (PDB: 9DW2) with white cartoon and transparent white surface.

https://doi.org/10.1038/s41557-026-02125-6



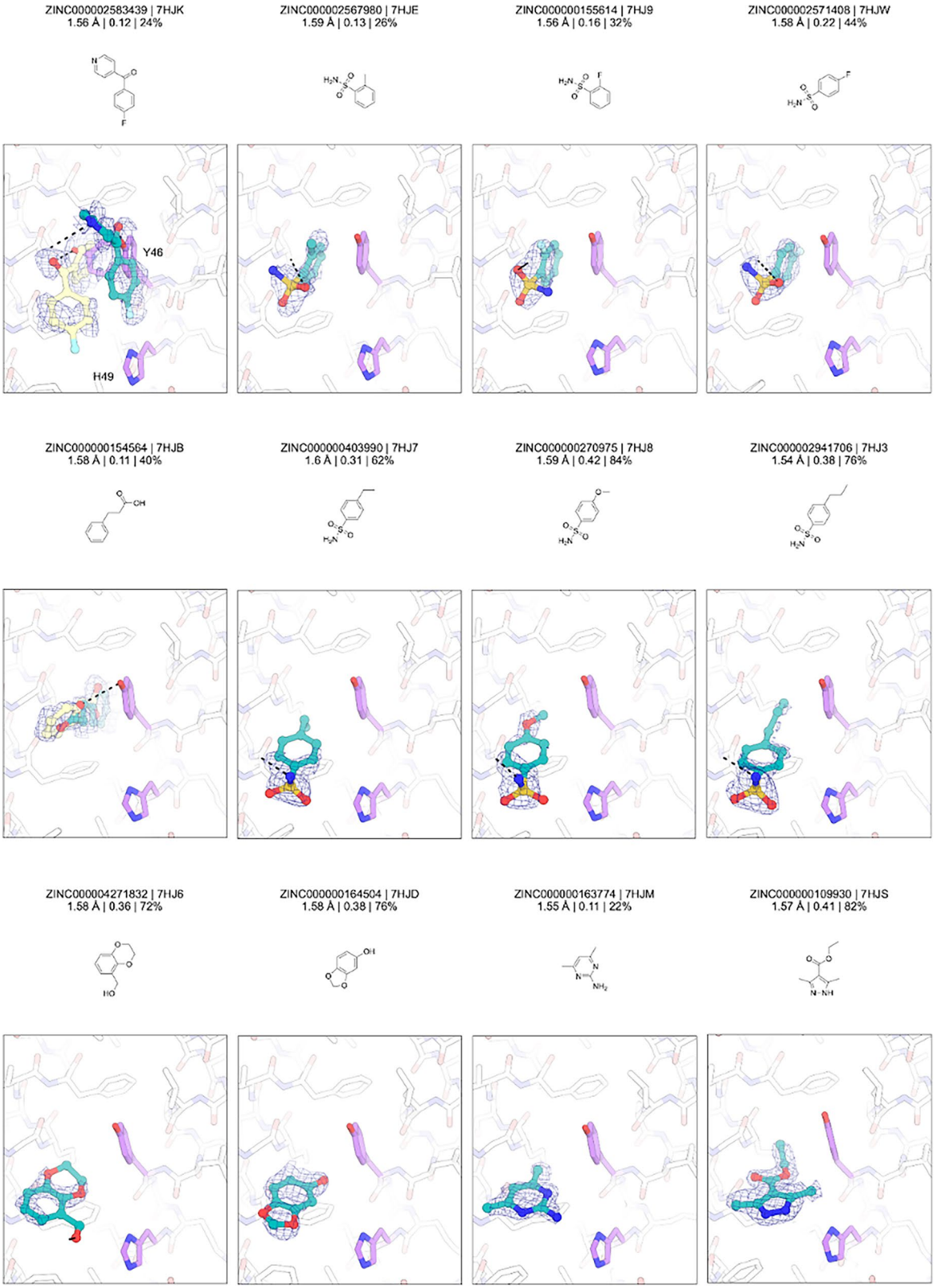
Extended Data Fig. 2 | Chemical structures and electron density maps for the ABLE-fragment complex structures (PDBs: 7HJA, 7HJ2, 7HJL, 7HJY, 7HK3, 7HK1, 7HK2, 7HJH, 7HIY, 7HIZ, 7HJV, 7HJC. PanDDA event maps (blue mesh, 2 σ) are contoured around fragments (teal/yellow sticks). The sidechains of residues Tyr46 and His49 are shown with purple sticks. Fragment names, PDB codes, resolution, 1-BDC value and refined occupancies are indicated. Hydrogen bonds are shown with dashed black lines. For clarity, residues 10–23 and 105–118 are hidden for fragments binding in the ABLE core. His49-mediated polar interaction could be found in the majority of fragments at site B.

https://doi.org/10.1038/s41557-026-02125-6



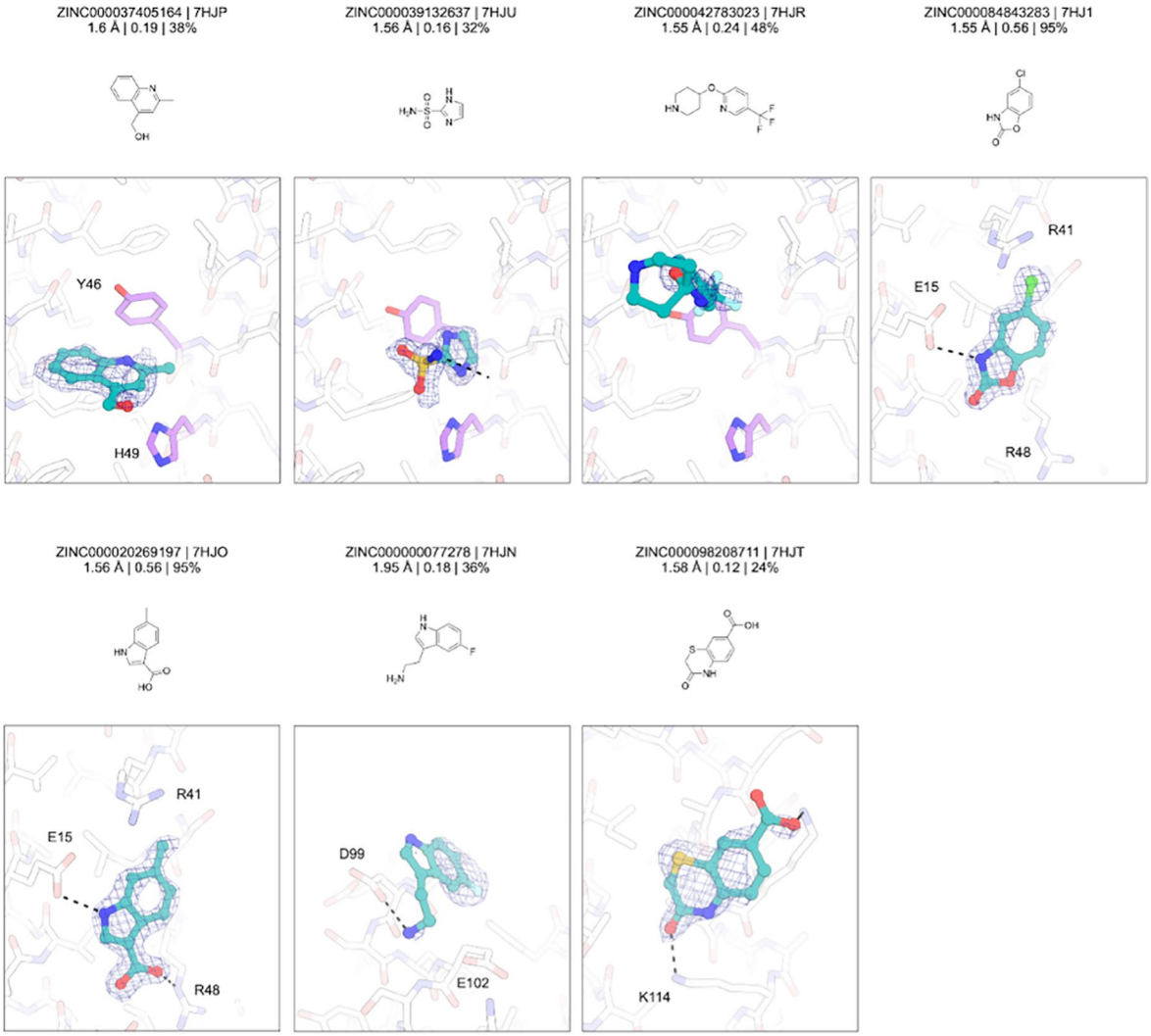
Extended Data Fig. 3 | Chemical structures and electron density maps for the ABLE-fragment complex structures (PDBs: 7HJI, 7HJ0, 7HJQ, 7HJ4, 7HJ5, 7HJZ, 7HK0, 7HK4, 7HJG, 7HJJ, 7HJX, 7HJF). PanDDA event maps (blue mesh, 2 σ) are contoured around fragments (teal/yellow sticks). The sidechains of residues Tyr46 and His49 are shown with purple sticks. Fragment names, PDB codes, resolution, 1-BDC value and refined occupancies are indicated. Hydrogen bonds are shown with dashed black lines. For clarity, residues 10–23 and 105–118 are hidden for fragments binding in the ABLE core. His49-mediated polar interaction could be found in the majority of fragments at site B.

https://doi.org/10.1038/s41557-026-02125-6



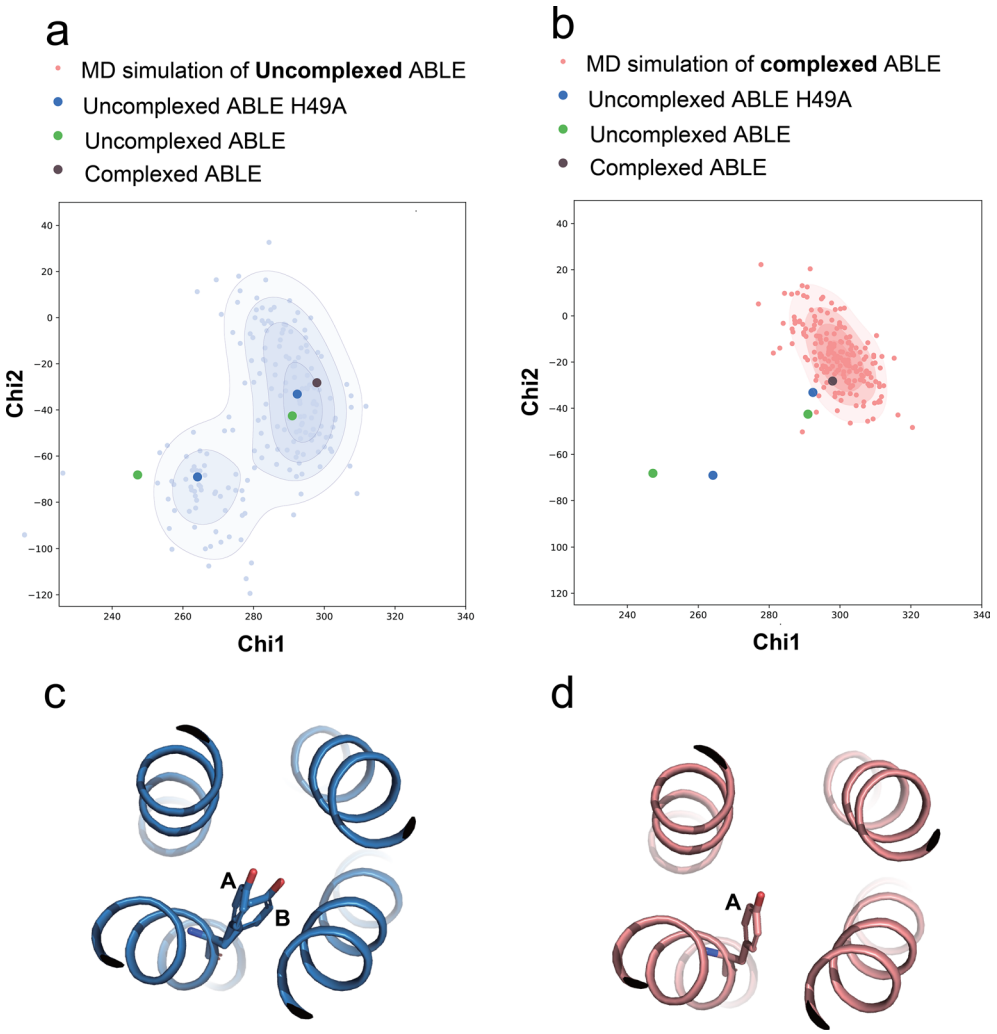
Extended Data Fig. 4 | Chemical structures and electron density maps for the ABLE-fragment complex structures (PDBs: 7HJK, 7HJE, 7HJ9, 7HJW, 7HJB, 7HJ7, 7HJ8, 7HJ3, 7HJ6, 7HJD, 7HJM, 7HJS). PanDDA event maps (blue mesh, 2 σ) are contoured around fragments (teal/yellow sticks). The sidechains of residues Tyr46 and His49 are shown with purple sticks. Fragment names, PDB codes, resolution, 1-BDC value and refined occupancies are indicated. Hydrogen bonds are shown with dashed black lines. For clarity, residues 10–23 and 105–118 are hidden for fragments binding in the ABLE core. His49-mediated polar interaction could be found in the majority of fragments at site B.

https://doi.org/10.1038/s41557-026-02125-6



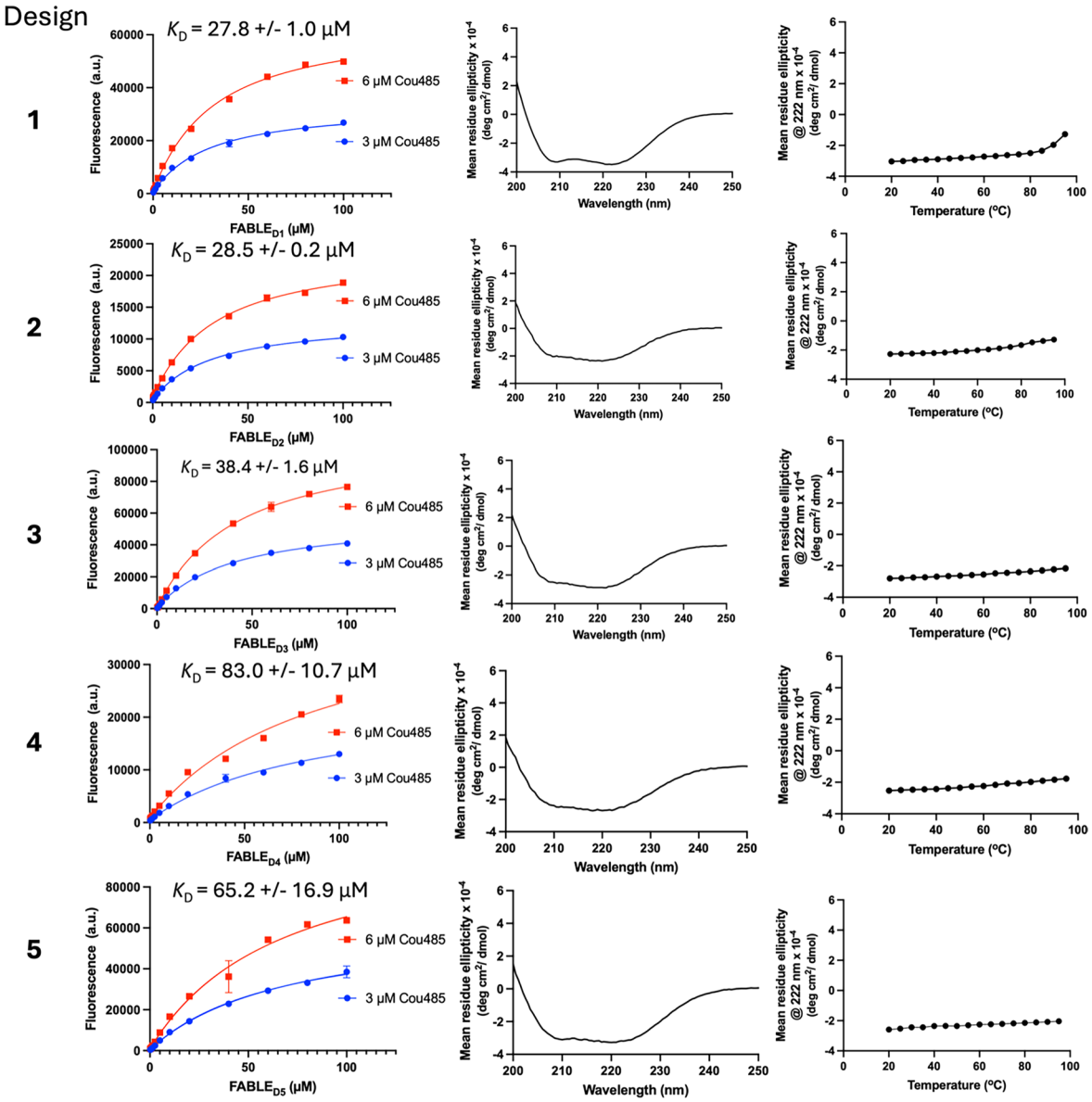
Extended Data Fig. 5 | Chemical structures and electron density maps for the ABLE-fragment complex structures (PDBs: 7HJP,7HJU,7HJR, 7HJ1,7HJO,7HJN,7HJT). PanDDA event maps (blue mesh, 2 σ) are contoured around fragments (teal/yellow sticks). The sidechains of residues Tyr46 and His49 are shown with purple sticks. Fragment names, PDB codes, resolution, 1-BDC value and refined occupancies are indicated. Hydrogen bonds are shown with dashed black lines. For clarity, residues 10–23 and 105–118 are hidden for fragments binding in the ABLE core. His49-mediated polar interaction could be found in the majority of fragments at site B.

https://doi.org/10.1038/s41557-026-02125-6



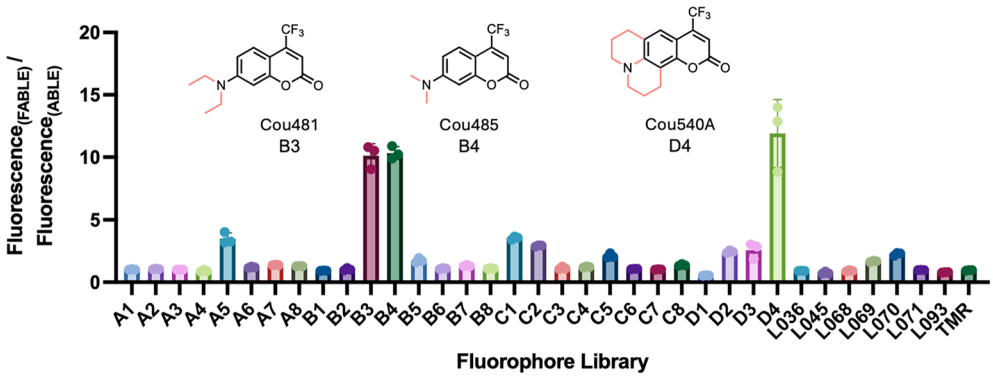
Extended Data Fig. 6 | Contour map of Tyr46 side chain conformation of ABLE from molecular dynamics. (**a**-**b**) Dynamics of uncomplexed ABLE (PDB: 6W6X) and ABLE-apixaban (PDB: 6W70) were explored by doing molecular simulation at 278 K for 500 ns with Amber. Chi1 and Chi2 from each state of these simulations and reported crystal structures of ABLEs (6W6X: uncompleted ABLE; 6W70: ABLE-Apixaban Complex; 6X8N: uncomplexed ABLE His49Ala) were extracted for plotting. Contours were generated from the frames of uncomplexed ABLE (**a**) and ABLE-apixaban complex contours (**b**) using an in-house script that utilizes the Gaussian kernel density estimate methods from scipy.stats module of SciPy Python Packages. (**c**-**d**) **A** and **B** conformations of Tyr46 sidechain were present in the uncomplexed ABLE crystal structure (**c**, PDB: 6W6X), while only A conformation of Tyr46 sidechain conformation was present at ABLE-apixaban complex (**d**, PDB: 6W70).

https://doi.org/10.1038/s41557-026-02125-6



Extended Data Fig. 7 | Characterization of five FABLE Designs. Left column: Fluorescence of Cou485 at two fixed concentrations of 3 µM and 6 µM were measured in the presence of increasing amounts of each protein. The dissociation constant was obtained by globally fitting a single-site binding model to data, as described at Supplementary Methods. The error bars represent standard deviations of three independent measurements. Middle Column: circular dichroism spectra shows that all the designs are helical proteins. Right column: temperature-dependent circular dichroism signals measured at 222 nm show that all designs are thermostable. The design 1 was designated as FABLE.

https://doi.org/10.1038/s41557-026-02125-6



Extended Data Fig. 8 | Exploring the chemical space of FABLE. The excitation and emission of fluorophores are in Supplementary Table 3. The error bars represent standard deviations of three measurements.

https://doi.org/10.1038/s41557-026-02125-6

Extended Data Table 1 | Kinetic parameters for KABLEs

Protein	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	pK_a	$(k_{cat}/K_M)_{max}$ ($M^{-1}s^{-1}$)
KABLE1	0.09 ± 0.001	0.20 ± 0.01	460 ± 20	8.8 ± 0.6	3,800 ± 1,400
KABLE1.4	99 ± 5	0.43 ± 0.05	230,000 ± 30,000	8.4 ± 0.1	630,000 ± 20,000
KABLE2.5[b]	550 ± 10	0.21 ± 0.01	2,500,000 ± 200,000	7.3 ± 0.1	3,200,000 ± 80,000

[a]Activities measured by plate reader at pH 8.0 buffer (20 mM TRIS, pH 8.0, 100 mM NaCl, 1.5% Acetonitrile), otherwise specified. Error bars represent the standard errors of the mean from at least three independent measurements. [b]Activities measured by stopped-flow to obtain high-quality kinetic data. The kinetic parameters from plate reader and UV-Vis spectrophotometer measurements were included in Supplementary Table 5.

https://doi.org/10.1038/s41557-026-02125-6

Extended Data Table 2 | Kinetic parameters of reported proteins catalyzing the Kemp elimination of 5-nitrobenzisoxazole with the base-mediated mechanism

Protein	Description	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (mM)	Amino acid efficiency[a]	Ref
34E4[c]	Catalytic antibody	5500	0.66	1.2	26	116
BSA[d]	Natural lipid carrier	6500	6.02	ND	11	117
KE07	Computational redesign	12.2	0.018	1.4	0.05	44
KE59	Computational redesign	163	0.29	1.8	0.66	44
KE70	Computational redesign	78	0.16	2.1	0.31	44
KE07 R7 10/11G	7 round DE of KE07	2590	1.37	0.54	10	57
KE59 R13_3/11H	13 round DE of KE59	60,430	9.53	0.16	240	89
KE70 R8 15/11E[e]	8 round DE of KE70	34,900	5.3	0.15	140	118
HG3	Computational redesign and rational engineering	425	0.68	1.6	1.4	47
HG3.17	17 round DE of HG3	230,000	700	3	760	45
HG3.R5	5 round DE of HG3	170,000	702	4.8	560	46
HG4	Ensemble-based redesign of HG3	103,000	ND	ND	340	50
HG649	Ensemble-based redesign of HG3	32,000	ND	ND	110	104
AlleyCat	Single mutation of Calmodulin	5.8	ND	ND	0.08	48
AlleyCat10	NMR-guided DE of AlleyCat7	4378	21.2	4.8	59	49
Lysozyme L99A/M102H[f]	Double mutation of Lysozyme	1.8	ND	ND	0.01	119
GNCA4 W229D F290W	Double mutation of GNCA4	5497	10	ND	21	51
V4[g]	DE of GNCA4 W229D F290W	200,000	635	3.14	710	120
V4-4	NMR-guided DE of V4	430,000	1700	3.9	1500	74
KE15	Computational redesign	35	0.022	0.63	0.14	44
KE15 Tyr167Lys+ R4[h]	Electrical field-guided design of KE15	403	0.31	0.77	1.6	121
tKSI	Natural enzyme	2.5	ND	ND	0.02	52
D38N tKSI	Single mutation of tKSI	17,000	ND	ND	140	52
Des27	Computational design	130	0.07	0.5	0.52	61
Des27.7	Computational redesign	12,700	2.85	0.21	51	61
F113L Des27.7	Single mutation on Des27.7	123,000	30	ND	490	61
KABLE1	*De Novo* Design	460; 3,800[b]	0.09	0.20	4; 33[b]	This work
KABLE1.4	Quadruple mutant of KABLE1	230,000; 630,000[b]	99	0.43	2,000; 5,400[b]	This work
KABLE2.5	NMR-guided DE of KABLE1.4	2,500,000; 3,200,000[b]	550	0.21	19,000; 27,000[b]	This work

[a]Amino acid catalytic efficiency, which is the quotient of catalytic efficiency (k_{cat}/K_M, $M^{-1}s^{-1}$), by the total number of amino acid for each protein. Kinetic parameters of KABLEs were measured at pH 8 buffer, unless specified. [b]$(k_{cat}/K_M)_{max}$ obtained from pH activity profile. [c]Data from ref. 118. [d]Data from ref. 119. [e]Data from ref. 120. [f]Data from ref. 121. [g]Data from ref. 122. [h]Data from ref. 123.

nature portfolio

Corresponding author(s): William F. DeGrado

Last updated by author(s): Dec 13, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	X-ray diffraction data was collected at beamline 8.3.1 of the Advanced Light Source (ALS) and beamlines 12-1 and 12-2 of the Stanford Synchrotron Radiation Lightsource (SSRL). Fluorescence data and enzyme activity data were taken using BioTek Synergy Neo2 Reader. Protein design was carried out with the following tools: Rosetta Modeling Suit 2020.08.61146 (https://www.rosettacommons.org/) LigandMPNN (https://github.com/dauparas/LigandMPNN) RaptorX-Single-ESM1b-ESM1v (https://github.com/AndersJing/RaptorX-Single) ESMFold with ESM-2 language model. (https://github.com/facebookresearch/esm) OmegaFold (https://github.com/HeliXonProtein/OmegaFold)
Data analysis	Diffraction images were indexed, integrated, and scaled with XDS and merged with Aimless. The initial model was analyzed with phenix.refine (version 1.21.1-5286) and COOT. Fragment binding was detected using the PanDDA algorithm packaged in CCP4 version 7.0. Molecular dynamic data were analyzed using VMD 1.9.4. Numerical data were analyzed with Microsoft Excel and GraphPad Prism 9.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Crystal structures of ABLE-fragment complexes have been deposited in the PDB with accession codes 7HIY, 7HIZ, 7HJ0, 7HJ1, 7HJ2, 7HJ3, 7HJ4, 7HJ5, 7HJ6, 7HJ7, 7HJ8, 7HJ9, 7HJA, 7HJB, 7HJC, 7HJD, 7HJE, 7HJF, 7HJG, 7HJH, 7HJI, 7HJJ, 7HJK, 7HJL, 7HJM, 7HJN, 7HJO, 7HJP, 7HJQ, 7HJR, 7HJS, 7HJT, 7HJU, 7HJV, 7HJW, 7HJX, 7HJY, 7HJZ, 7HK0, 7HK1, 7HK2, 7HK3 and 7HK4. Summary of data collection, modeling and refinement of ABLE-fragment complexes is at Supplemental File 1. Structure factor intensities (unmerged, merged, and merged/scaled), PanDDA input and output files including Z-map and event maps in CCP4 format, and refined models including the fragment-bound state extracted from multi-state models were uploaded to Zenodo (DOI: 10.5281/zenodo.13913848). Crystal structures of FABLE and KABLE have also been deposited in the PDB with accession codes 9DWA, 9DWB, 9DWC, 9NOI, 9N0J. X-ray data collection and refinement statistics for ABLE / FABLE / KABLE structures are at Supplemental File 1. Scripts used in design and analysis have been deposited at Zenodo with open access (DOI:10.5281/zenodo.17935960). All other data not included in the main text and supplemental files could be shared upon request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We mounted 447 crystals soaked with fragments and were able to obtain high quality datasets from 321. The entire Enamine Essential fragment library was used without prior selection. The total number of unique fragments with high-quality datasets was 242. The number of designs ordered for experimental validation was estimated on confidence of computational results and experimental work load.
Data exclusions	No data exclusion was done.
Replication	All kinetic data of Kemp eliminase were run in triplicate with at least two batches of protein expression. The activity of KABLE1, KABLE1.3, and KABLE1.4 were tested by two individuals from the initial step of plasmid transformation to the final step of the kinetic assay. Fluorescence data of FABLE were measured in triplicate with at least two batches of protein expression. All repeats result in consistent activity results.
Randomization	Randomization was not applied in this study, due to the independent relation of results with sample application.
Blinding	Blinding was used when testing the activity of designed Kemp eliminases. The person conducting experiment test was unaware of sequence identities of the designs.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*