

# Engineering the protein dynamics of an ancestral luciferase

Andrea Schenk Mayerova <sup>1,2,6</sup>, Gaspar P. Pinto<sup>1,2,6</sup>, Martin Toul <sup>1,2,6</sup>, Martin Marek <sup>2,6</sup>, Lenka Hernychova <sup>3</sup>, Joan Planas-Iglesias <sup>1,2</sup>, Veronika Daniel Liskova<sup>1,2</sup>, Daniel Pluskal<sup>2</sup>, Michal Vasina<sup>1,2</sup>, Stephane Emond <sup>4</sup>, Mark Dörr<sup>5</sup>, Radka Chaloupkova<sup>2</sup>, David Bednar<sup>1,2</sup>, Zbynek Prokop<sup>1,2</sup>, Florian Hollfelder <sup>4</sup>✉, Uwe T. Bornscheuer <sup>5</sup>✉ & Jiri Damborsky <sup>1,2</sup>✉

Protein dynamics are often invoked in explanations of enzyme catalysis, but their design has proven elusive. Here we track the role of dynamics in evolution, starting from the evolvable and thermostable ancestral protein Anc<sup>HLD-RLuc</sup> which catalyses both dehalogenase and luciferase reactions. Insertion-deletion (InDel) backbone mutagenesis of Anc<sup>HLD-RLuc</sup> challenged the scaffold dynamics. Screening for both activities reveals InDel mutations localized in three distinct regions that lead to altered protein dynamics (based on crystallographic B-factors, hydrogen exchange, and molecular dynamics simulations). An anisotropic network model highlights the importance of the conformational flexibility of a loop-helix fragment of *Renilla* luciferases for ligand binding. Transplantation of this dynamic fragment leads to lower product inhibition and highly stable glow-type bioluminescence. The success of our approach suggests that a strategy comprising (i) constructing a stable and evolvable template, (ii) mapping functional regions by backbone mutagenesis, and (iii) transplantation of dynamic features, can lead to functionally innovative proteins.

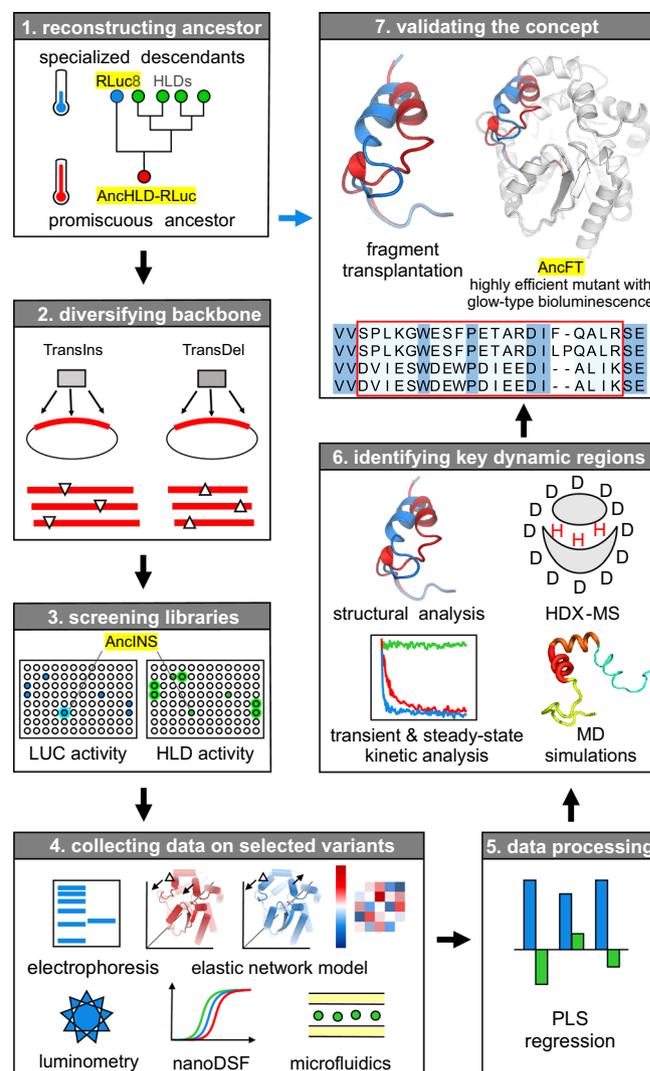
<sup>1</sup>International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic. <sup>2</sup>Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic. <sup>3</sup>Research Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, Brno, Czech Republic. <sup>4</sup>Department of Biochemistry, University of Cambridge, Cambridge, UK. <sup>5</sup>Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, Greifswald, Germany. <sup>6</sup>These authors contributed equally: Andrea Schenk Mayerova, Gaspar P. Pinto, Martin Toul, Martin Marek. ✉email: [fh111@cam.ac.uk](mailto:fh111@cam.ac.uk); [uwe.bornscheuer@uni-greifswald.de](mailto:uwe.bornscheuer@uni-greifswald.de); [jiri@chemi.muni.cz](mailto:jiri@chemi.muni.cz)

Contemporary biocatalysts may originate from a small number of possibly multifunctional common ancestors and a limited number of structural folds. Natural enzymes have undergone billions of years of evolution to generate vast functional diversity and strikingly precise and efficient activities. This diversity resulted from complex evolutionary processes<sup>1</sup>, which have been grouped into two complementary mechanisms: creeping and leaping evolution<sup>2</sup>. The former involves relatively minor functional changes that generally increase specificity or activity, whereas the latter involves radical shifts that introduce functional innovations such as the ability to bind a completely different substrate or change an enzyme's mechanism. While small adaptive changes may result from point substitutions introduced during evolution<sup>3</sup>, larger functional leaps may require more profound rearrangements of the protein backbone<sup>4</sup>. However, the potential for innovation comes at the price of disruption and destabilization. One way to harvest the effects of profound modifications is to infer ancestral enzymes<sup>5</sup> with features that enhance evolvability: stability and promiscuity<sup>6,7</sup>. This enables the creation of robust generalist scaffolds that may be more catalytically versatile because they are less burdened by adaptive pressure towards a specific function than their modern-day counterparts<sup>8,9</sup>. If ancestors are indeed more stable, their capacity to buffer the large mutational load arising from backbone modifications<sup>4</sup> may be enhanced.

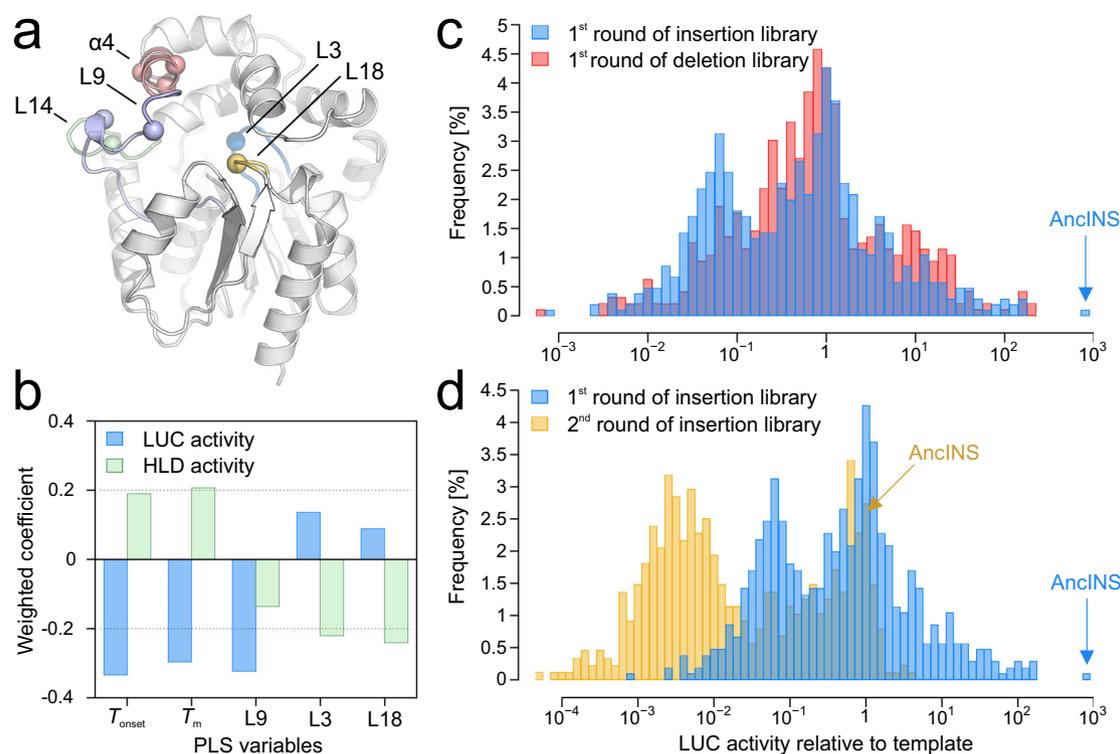
Most directed evolution efforts rely on point substitutions. There have been far fewer reports of experimental insertions or deletions (InDels), despite their relatively frequent and beneficial occurrence in natural evolution<sup>10</sup>. Various genomic analyses show that the ratio of InDels to point substitutions in protein-coding regions typically ranges from 1:5 in primates to 1:20 in bacteria, indicating that InDels are typically subjected to stronger purifying selection than substitutions during evolution. This is due to their potentially more deleterious effects, including losses of stability, disruptions of secondary structure elements and/or perturbations of folding pathways<sup>11</sup>. To explore the potential effects of InDel mutagenesis on functional proteins, we previously developed TRIAD (transposition-based random insertions and deletions)<sup>12,13</sup>, a method for generating random InDel libraries that provides ready access to variants that cannot be obtained by substitution mutagenesis <https://doi.org/10.21203/rs.3.rs.1448/v1>. TRIAD is transposon-based and was shown to have only minimal sequence bias, with >85% of all possible sites shown to be targeted by the transposon<sup>12</sup>. In the present work, TRIAD is applied to a particularly stable ancestral protein<sup>14</sup>, the recently designed and characterized Anc<sup>HLD-RLuc</sup> that was reconstructed from the catalytically distinct but evolutionarily and structurally related haloalkane dehalogenases<sup>15</sup> (HLD, EC 3.8.1.5) and *Renilla* luciferase<sup>16</sup> (RLuc, EC 1.13.12.5). This ancestor<sup>14</sup> is bifunctional and catalytically versatile, with a dehalogenase activity comparable to contemporary HLDs and a promiscuous luciferase activity almost 7000-fold lower than that of the stabilized RLuc8, a popular molecular probe. The light-producing reaction catalysed by the *Renilla*-type luciferase is one of the most widely used biochemical reactions in molecular and cell biology research. Based on conformational differences of the RLuc8 backbone in crystal structures<sup>17</sup>, it had been suggested that the opening of the L9 loop might be relevant for luciferase substrate binding. The availability of Anc<sup>HLD-RLuc</sup>, a stable bi-functional scaffold<sup>14</sup>, enabled investigation of structure-function relationships of luciferase activity.

Here we present a three-step protein engineering strategy designed to exploit the effects of InDels as promoters of evolutionary innovation: (i) constructing a robust and evolvable template, (ii) mapping functional regions by backbone mutagenesis and multivariate statistics and (iii) transplantation of a dynamic structural (loop-helix) feature. Its application to the bifunctional

protein Anc<sup>HLD-RLuc</sup> (Fig. 1) results in an engineered 7000-fold more efficient catalyst with 100-fold longer glow-type bioluminescence applicable as a molecular probe for use in bacterial as well as mammalian cells.



**Fig. 1 | Illustration of the strategy for semi-rational engineering of protein dynamics.** Exploratory phase—**1** A thermostable ancestral protein, Anc<sup>HLD-RLuc</sup>, provided a robust and evolvable template that can withstand the destabilizing effects of protein backbone engineering. **2** Libraries of single triplet insertion and deletion variants were created following the TRIAD method based on the use of engineered transposons (TransIns and TransDel)<sup>12,13</sup>. **3** Screening of the libraries led to identification of the improved insertion mutant AncINS. **4** Twenty-five mutants with significant changes in luciferase (LUC) and haloalkane dehalogenase (HLD) activities were characterized using bioinformatics, microscale techniques (nano differential scanning fluorimetry—nanoDSF), and microfluidics. **5** Structure-function relationships were described employing partial least squares (PLS) multivariate statistics. **6** Dynamic elements required for efficient catalysis were identified by structural, kinetic, biophysical and computational characterization (molecular dynamics (MD) simulations) of AncINS. Validation phase—**7** Knowledge obtained during the exploratory phase was validated by transplanting a relevant dynamic fragment from the specialized descendant into the ancestor, yielding an enzyme, Anc<sup>FT</sup>, with 7000-fold higher catalytic efficiency than Anc<sup>HLD-RLuc</sup> and 100-fold longer glow-type bioluminescence than the flash-type *Renilla* luciferase RLuc8. Key mutants discussed in this study are highlighted in yellow.



**Fig. 2 Quantitative structure-function relationships and analysis of the effects of template and mutation type on activity.** **a** Crystal structure of the thermostable Anc<sup>HLD-RLuc</sup> (PDB ID 6G75) showing the positions of the L9 loop (light blue), the  $\alpha 4$  helix (salmon) and the L14 loop (pale green), where InDels (spheres) resulted in an increase of luciferase (LUC) activity. Correlated motions of L3 (marine) and L18 (yellow) loops carrying two of the five catalytic amino acids (spheres) were identified as significant contributors to dehalogenase (HLD) activity by the partial least squares (PLS) regression analysis. **b** Weighted coefficients quantifying contributions of variables indicated in the PLS models to explain the variance in HLD (green) and LUC (blue) activity. Note that different directionality of these coefficients for all variables, except loop L9, suggests different mechanistic and structural requirements for the two enzymatic functions studied. PLS generated models that explained substantial amounts of the variation in both LUC activity ( $R^2 = 0.73$ ,  $Q^2 = 0.67$ ,  $n = 25$  is the number of InDel variants) and HLD activity ( $R^2 = 0.63$ ,  $Q^2 = 0.54$ ,  $n = 25$  is the number of InDel variants). To obtain  $Q^2$  the model was recalculated 999 times with a randomly re-ordered dependent variables. **c** Comparison of frequencies of variants with their activities relative to the ancestral protein in the first-round insertion (R1I, blue) and deletion (R1D, red) libraries. The template for R1I and R1D was Anc<sup>HLD-RLuc</sup>. The best insertion variant AncINS is indicated by an arrow. **d** Comparison of frequencies of variants with indicated activities relative to their respective templates observed in the first round of insertion library (R1I, blue) and the second round of insertion library (R2I, yellow). Note that the starting template Anc<sup>HLD-RLuc</sup> for R1I and R1D is less active than the starting template AncINS for R2I. The template for R2I was AncINS. Source data is available as a Source data file for Fig. 2.

## Results

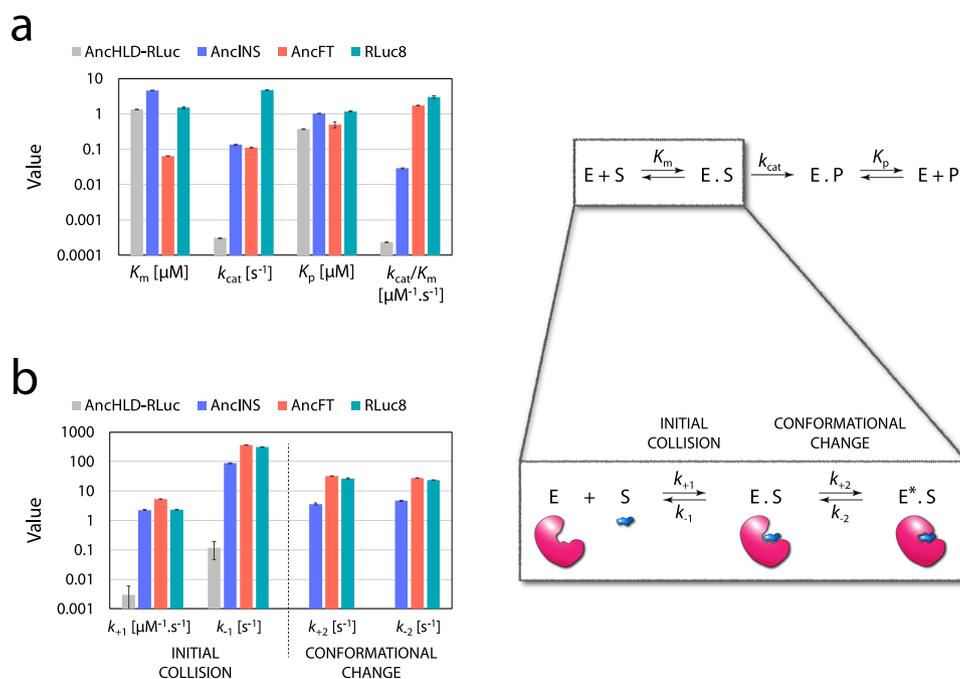
### Structure–function relationships by analysis of InDel libraries.

Libraries of Anc<sup>HLD-RLuc</sup> with random single amino acid insertions (1st round, R1I) and deletions (1st round, R1D) distributed over the length of the protein were constructed using TRIAD<sup>12</sup> and screened for LUC and HLD activities (Supplementary Note 1). All hits with significantly improved LUC activity had backbone alterations localized in one of three regions of the enzyme cap domain: the L9 loop, the  $\alpha 4$  helix or the L14 loop (Fig. 2a, Supplementary Table 1). We selected 11 variants (10 from R1I, 1 from R1D) for further analysis and complemented them with 15 variants with low LUC activity that retained HLD activity (8 from R1D and 7 from R1I, Source data file). All variants were expressed, purified and subjected to activity and stability measurements and thermodynamic analyses using a microfluidic approach (Supplementary Fig. 1). To understand the role of cooperativity and long-range interactions after mutagenesis, we used an anisotropic network model (ANM, Supplementary Note 2) to calculate the cross-correlation of motions of selected structural fragments (Supplementary Tables 2, 3, Supplementary Note 3) to the regions carrying the InDel mutations of the selected variants.

Multivariate partial least squares (PLS) regression (Supplementary Note 4) revealed that the initiation of protein unfolding ( $T_{\text{onset}}$ )

and its midpoint ( $T_m$ ) were the strongest predictors for both LUC and HLD activities. At  $T_{\text{onset}}$  (0.5% deviation from a linear fit on the 330 nm/350 nm vs. temperature curve) a protein starts to unfold, while at  $T_m$  (inflection point on the 330 nm/350 nm vs. temperature curve) 50% of the protein is unfolded. The directionality of the relationships was the opposite for HLD and LUC activities (Fig. 2b). This suggested that stabilities are linked, positively and negatively, to the two activities with their distinct chemical mechanisms and structural requirements. Moreover, cross-correlated motions of the mutated positions to the L9 loop were identified as very strong contributors to LUC activity, and cross-correlated motions to the L3 and L18 loops (which encompass catalytic residues<sup>14,18</sup>) as key contributors to HLD activity.

All the positive hits had significantly reduced thermal stability compared to the ancestral protein (up to  $-20^\circ\text{C}$  lower  $T_{\text{onset}}$  and  $T_m$  values; Source data file), in agreement with the notion that a stable template is crucial for obtaining functional mutants. Moreover, we assumed that using a highly stable starting point for engineering would promote evolvability by allowing the protein to accept a wider range of mutations while retaining its native fold<sup>19</sup>, thereby enabling multiple rounds of directed evolution. To probe this hypothesis, the best variant from R1I library (AncINS) carrying an insertion and substitution in the  $\alpha 4$  helix was used as the template for construction of another amino



**Fig. 3 Steady-state and transient kinetic analysis of CTZ conversion.** The kinetic models consist of an enzyme E, a substrate S, an enzyme-substrate complex in two conformations (E.S and E\*.S), an enzyme-product complex E.P, and a product P. **a** Steady-state kinetic parameters (Michaelis constant  $K_m$ , turnover number  $k_{cat}$ , enzyme-product complex dissociation constant  $K_p$ ) were determined with the substrate CTZ in 100 mM phosphate buffer at pH 7.5 and 37 °C by global analysis of triplicates of full progress curves recorded with at least five concentrations of CTZ. **b** Results of pre-steady-state kinetic analysis of the CTZ substrate binding in 100 mM phosphate buffer at pH 7.5 and a lower temperature (15 °C). This enabled identification of the induced fit substrate binding mechanism, involving initial collision of the enzyme and the substrate (described by forward rate constant  $k_{+1}$  and reverse rate constant  $k_{-1}$ ), followed by a conformational change of the enzyme induced by the bound substrate (described by forward rate constant  $k_{+2}$  and reverse rate constant  $k_{-2}$ ). A simple binding mechanism including only the first step was observed for Anc<sup>HLD</sup>-RLuc. The kinetic parameters were determined by global fitting of tryptophan fluorescence traces obtained with at least 10 concentrations of CTZ and 10 concentrations of each tested enzyme, in each case with seven replicates (Supplementary Note 6). The data are presented as best fit values  $\pm$  standard errors (S.E.) calculated from the covariance matrix during nonlinear regression. Source data is available as a Source data file for Fig. 3.

acid insertion library (2nd round, R2I) that was screened for further improvement in LUC activity. Statistical analysis of activity data shows that R1I and R1D yielded populations of variants with similar LUC activity patterns, and most of the increases in relative activity obtained in these rounds were compromised in R2I (Fig. 2c, d). The proportions of variants showing at least doubled activity of the starting template generated in R1I, R1D and R2I were 13%, 15.5% and just 0.5%, respectively. As no significant activity improvements were obtained in the second round of insertion mutagenesis, no further rounds were pursued.

The finding that a single round of InDel mutagenesis of the stabilized template already provided mutants with >100-fold increases in LUC activity raises the question: why are the structural elements (L9 and L14 loops and  $\alpha$ 4 helix), highlighted by directed evolution experiments, crucial for LUC activity? To address this question, the mutant AncINS, carrying an insertion in the  $\alpha$ 4 helix, was selected as a representative and studied by transient and steady-state kinetics.

**Kinetic analysis underlines the importance of conformational flexibility for substrate binding.** To gain insight into the catalytic mechanism, steady-state kinetics of LUC activity of the three luciferases (AncINS, the template Anc<sup>HLD</sup>-RLuc and the modern variant RLuc8; Fig. 1) were measured with the substrate coelenterazine (CTZ). Complete luminescence progress curves (Supplementary Fig. 5) gave, by numerical simulation, estimates of the turnover number ( $k_{cat}$ ) and the enzyme specificity constant

( $k_{cat}/K_m$ ).  $k_{cat}$  provides a lower limit for each first-order rate constant following binding of a substrate through product release, while  $k_{cat}/K_m$  indicates a lower limit of the rate for substrate binding (Fig. 3a). The parameters obtained for RLuc8 correspond well with previously reported values derived from initial rate measurements combined with quantum yield calibration (Supplementary Table 4)<sup>20</sup>. Moreover, following a time course of the substrate-to-product conversion, our data provide a sensitive estimate of the equilibrium dissociation constants for each enzyme-product complex ( $K_p$ ) that cannot be obtained by conventional initial rate analysis. Both  $k_{cat}$  and  $k_{cat}/K_m$  values of AncINS were in-between those of Anc<sup>HLD</sup>-RLuc and RLuc8, indicating a 124-fold enhancement of its catalytic efficiency relative to that of the ancestral enzyme.

We also investigated by pre-steady-state kinetic analysis, whether conformational changes of the protein are essential for effective binding of the bulky CTZ substrate (Fig. 3b). The combination of analytical data processing and global fitting by numerical integration revealed a two-step induced fit substrate-binding mechanism: initial collision of the enzyme with the substrate, followed by an induced conformational change of the enzyme (Fig. 3b, Supplementary Fig. 6). Details and the rationale of the mechanistic analysis are described in the Supplementary Information (Supplementary Note 5). The kinetic constants are consistent with a model in which the reaction catalysed by Anc<sup>HLD</sup>-RLuc involves a slow simple binding mechanism, with no sign of conformational flexibility. By contrast, we detected conformational flexibility of AncINS, accompanied by 1000-fold

faster binding, based on determined  $k_{+1}$  and  $k_{-1}$ . The changes in AncINS increased rates of the initial collision step to match those of RLuc8, but a subsequent conformational change was still 10-fold slower. Faster kinetics of substrate binding after backbone modification in the  $\alpha 4$  helix suggest that this region plays a significant role in the initial step of the catalytic cycle.

**Backbone conformational differences in crystal structures.** To investigate the relationship between the structural changes induced in Anc<sup>HLD-RLuc</sup> by InDel mutagenesis and enhanced kinetics of substrate binding, we crystallized AncINS (PDB ID 6S6E, Supplementary Table 6) and compared its structure to previously published structures of Anc<sup>HLD-RLuc</sup> (PDB ID 6G75) and RLuc8 (PDB ID 2PSF). These proteins have identical  $\alpha/\beta$ -hydrolase folds but differ in their cap domains. Important changes were identified in the conformation of the  $\alpha 4$  helix and L9 loop in the cap domain, size of the active site cavity, width of the tunnel mouth, and active site accessibility (Fig. 4).

The asymmetric unit of the crystal lattice of AncINS contains two monomers (chains A and B). Chain A has a structure similar to the Anc<sup>HLD-RLuc</sup> template but features a  $\pi$ -helix bulge in the  $\alpha 4$  helix where the L162 insertion and F163P substitution occurred. In AncINS chain B, the  $\alpha 4$  helix is markedly distorted towards the  $\alpha 5$  helix. Moreover, the electron density map for the L9- $\alpha 4$  fragment is not perfectly resolved. Some side chains are poorly visible or invisible, suggesting that appreciable internal motion occurs in this region. RLuc8 also has two monomers in the asymmetric unit; chain B is similar to that of Anc<sup>HLD-RLuc</sup>, while chain A is in an open conformation with the  $\alpha 4$  helix pointing away from the  $\alpha 5$  helix.

In AncINS, the inserted L162 occupies the same position as the bulky F162 in Anc<sup>HLD-RLuc</sup>, making the active site cavity of AncINS bigger than that of Anc<sup>HLD-RLuc</sup>. The F163P substitution in AncINS disrupts the  $\alpha 4$  helix because proline acts as a helix breaker, introducing a kink into the polypeptide backbone. The resulting distortion closes the main access tunnel in the chain B of AncINS. In RLuc8, the situation is a little bit more complex. The sequence alignment (Supplementary Fig. 9) indicates that I163 is a conserved amino acid, but it does not occupy the same position as I161 in Anc<sup>HLD-RLuc</sup>. Instead, it is flipped down into the position occupied by the bulky F162 in Anc<sup>HLD-RLuc</sup>. The outward conformation of the  $\alpha 4$  helix in the cap domain of RLuc8 chain A gives it the largest active site cavity of the crystal structures studied.

To ascertain the role of the L9- $\alpha 4$  fragment in substrate binding, we tried to co-crystallize RLuc8 with CTZ. Despite our best efforts, we did not obtain sufficiently well diffracting crystals. In contrast, mixing the RLuc8-W121F/E144Q mutant with CTZ yielded a high-resolution complex structure (Fig. 5a; Supplementary Table 6) with coelenteramide (CEI), which is the product of the LUC reaction. Unlike in an RLuc8-CEI complex structure<sup>17</sup>, the CEI molecule in our complex is inverted by nearly 180°, which allows its accommodation in the luciferase active site cavity (Fig. 5b, Supplementary Fig. 10). The rest of the CEI molecule occupies the main enzyme access tunnel, where it is tightly wrapped by multiple, predominantly hydrophobic, residues: V146, I150, I159, V185, F181, F261, F262, H285 and I266.

Interestingly, many of the residues directly interacting with the CEI are located on the loop L9 (V146, I150, W156 and I159) and the loop L14 (I223 and P224), identified as important for catalysis by our directed evolution experiments. The  $\alpha 4$  helix is in an open conformation relative to the  $\alpha 5$  helix (Fig. 5c). Ultimately, the L9- $\alpha 4$  fragment directly affects the opening and closing of the access tunnels and we deduce that it is involved in substrate binding and product release. Thus, our structural data support conclusions

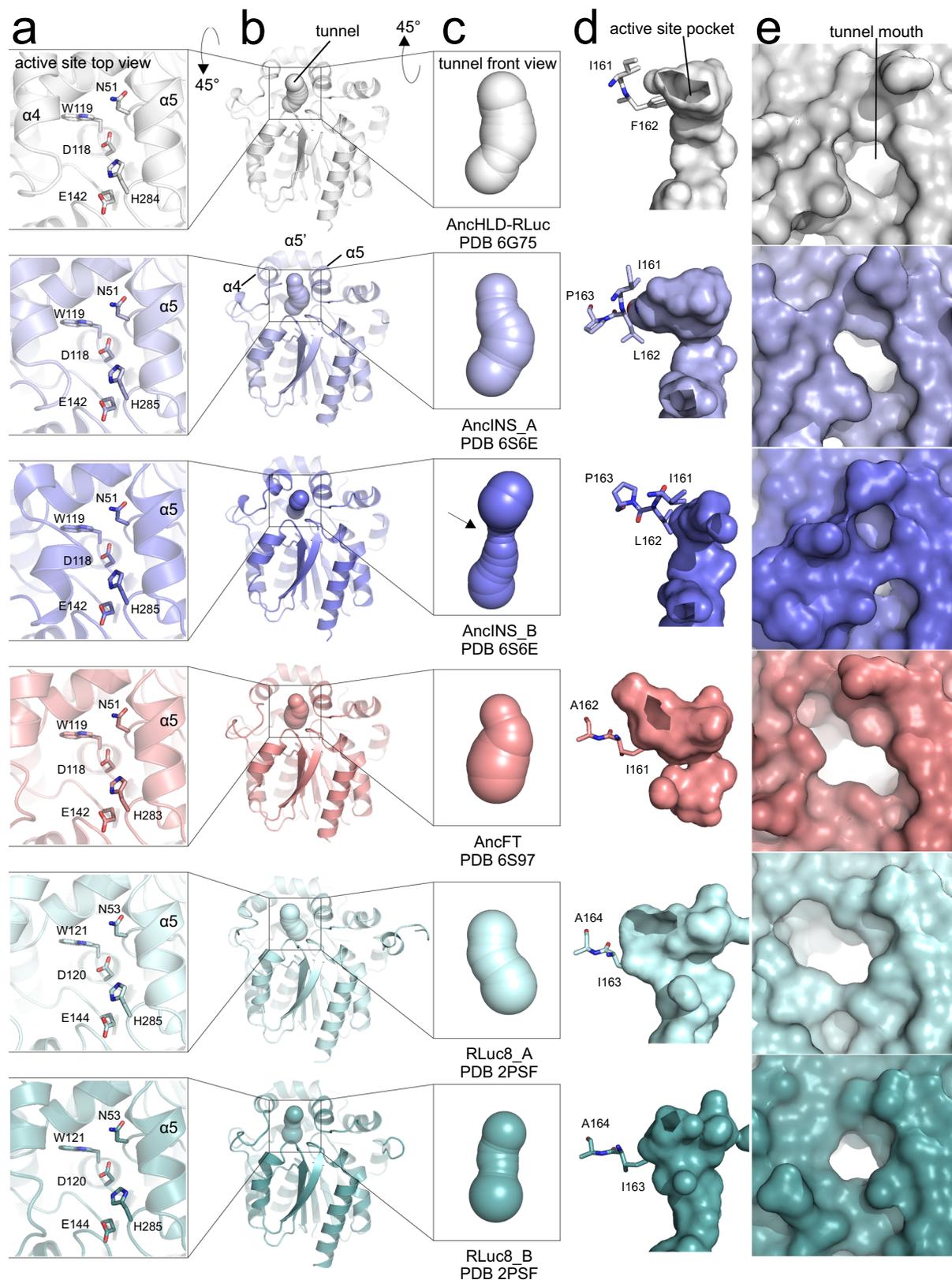
from InDel mutagenesis and kinetic studies that the L9- $\alpha 4$  fragment is the key hotspot region for the introduction of LUC activity.

**Profiling protein dynamics by mass spectrometry and molecular dynamics.** To analyse the effects of dynamics on LUC activity, we recorded hydrogen-deuterium exchange mass spectrometry (HDX-MS) time courses and complementary molecular dynamics (MD) simulations (Fig. 6). In this context, it should be emphasized that HDX-MS reveals changes occurring over seconds or minutes, whereas MD simulations cover a few microseconds (in this case 4.8  $\mu$ s). Kinetic analyses were carried out with the substrate, while HDX-MS and MD simulations were performed in the absence of any ligand. HDX-MS captures both dynamics and solvation.

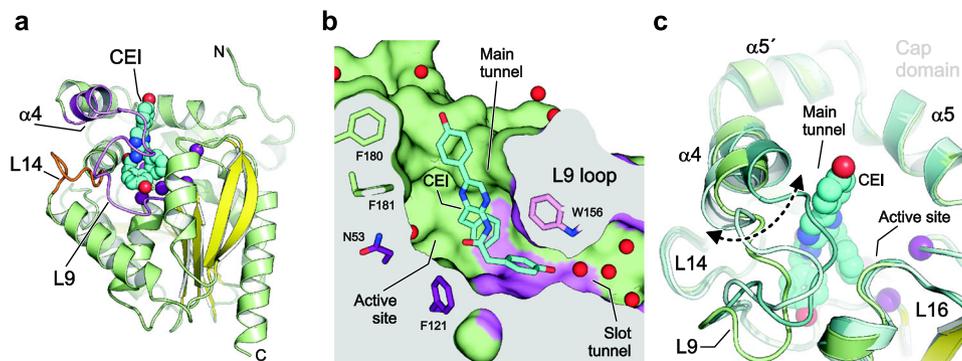
Backbone amide hydrogen-deuterium exchange is particularly sensitive to hydrogen bonding and thus reflects hydrogen bonding strength, conformational dynamics, and the solvent accessibility of protein structures. The HDX-MS profiles (Fig. 6c, Supplementary Fig. 11) show that AncINS was deuterated more rapidly overall than the template Anc<sup>HLD-RLuc</sup>, and that the modification of its  $\alpha 4$  helix backbone leads to greater solvation and more pronounced dynamics, particularly in the cap domain region (Supplementary Fig. 11). Moreover, AncINS exhibited very high overall deuteration within 10 and 60 s, with the amino acids between residues W151 (L9 loop) and F180 ( $\alpha 5$  helix) of the cap domain being most extensively deuterated. Residues S212-E235 (which comprise the L14 loop) were also appreciably more deuterated than their counterparts in Anc<sup>HLD-RLuc</sup>. The deuteration of the  $\alpha 4$ - $\alpha 5$  fragment in RLuc8 reached a deuteration level comparable to the level observed in AncINS within 60 s, but it increased further in 300 s. In addition, the L9 loop of RLuc8 was less extensively deuterated than the corresponding loop in Anc<sup>HLD-RLuc</sup> and the region S212-E235 (L14 loop) of RLuc8 was more heavily deuterated than in Anc<sup>HLD-RLuc</sup>. These deuteration patterns in HDX-MS indicate a possible correlation between the dynamics of hot spot regions comprising the  $\alpha 4$  and  $\alpha 5'$  helices and L14 loop and high LUC activity. The protein dynamics are coupled to efficient binding of the bulky substrate and release of the structurally similar product, based on the kinetic constants  $k_{+1}$  and  $k_{-1}$  obtained by the pre-steady state kinetics (Fig. 3b, Supplementary Fig. 6).

The regions encompassing the L9- $\alpha 4$  fragment and L14 loop are important structural elements that line the main access tunnels connecting the buried active site to the surrounding solvent. Their motions are concerted based on ANM and significantly affect the active site's volume, while fully preserving the catalytic residues' geometry. Since HDX-MS experiments measure the exchange of backbone amide hydrogens with deuterium, we used backbone B-factors from MD simulations to characterize the enzymes' dynamics<sup>14,21</sup>. The overall B-factors obtained for the cap domain of RLuc8 were highest in the  $\alpha 4$  helix, followed by the  $\alpha 5'$  helix and L9 loop (Fig. 6b). The B-factors for the L14 loop were similar to those of the L9 loop, but values for the rest of the cap domain were below average (Supplementary Fig. 12). The dynamic profile of the Anc<sup>HLD-RLuc</sup> template differs significantly: unlike in RLuc8, the dynamics are most pronounced in the L9 loop. Interestingly, the AncINS mutant exhibits markedly higher B-factors in the  $\alpha 4$  helix, making it the most dynamic region together with the L9 loop (Fig. 6b).

**Validation by fragment transplantation yields stable glow-type bioluminescence.** To verify the importance of the hotspot regions highlighted by biophysical analysis, we transplanted the sequence



**Fig. 4 Comparison of structural changes in the apo forms of the enzymes.** Anc<sup>HLD</sup>-RLuc (white), AncINS (chain A light blue, chain B slate), AncFT (salmon), RLuc8 (chain A pale cyan, chain B light teal). **a** Top view of the active site. The putative catalytic pentad and  $\alpha 5$  helix occupy the same position in all proteins, but the  $\alpha 4$  helix adopts different conformations. **b** Crystal structure of the whole protein, with the main access tunnel (identified using Cover 3.02<sup>23</sup>) shown as spheres. The outward/inward movement of the  $\alpha 4$  helix causes opening/closing of the tunnel. **c** Front view of the main tunnel. The black arrow indicates the visible constriction due to closure of the  $\alpha 4$  helix. **d** The size of the active site cavity is changed by repositioning of amino acids in the  $\alpha 4$  helix. **e** The size of the main tunnel mouth is changed by movements of the  $\alpha 4$  helix and L9 loop.



**Fig. 5** Structural characterization of coelenteramide binding to the catalytically defective RLuc8-W121F/E144Q mutant. **a** Cartoon representation of the overall structure of the RLuc8-W121F/E144Q mutant with coelenteramide (CEI) in its active site. CEI is the product of the LUC reaction and shown as cyan space-filling spheres. Residues of the conserved catalytic pentad are shown as purple spheres; the central eight-stranded  $\beta$ -sheet is coloured yellow; the  $\alpha 4$  helix and L9 loop (L9- $\alpha 4$  element) are coloured violet, and the L14 loop is coloured orange. **b** Cutaway surface representation of the enzyme active site cavity with the bound CEI (shown as cyan sticks). The colouring is the same as in panel **(a)**. Water molecules are shown as red spheres. **c** Close-up view of structural superposition of RLuc8-W121F/E144Q (green), RLuc8 (PDB ID 2PSF A; cyan) and RLuc8 (PDB ID 2PSF B; teal). CEI is shown as cyan space-filling spheres. Note the conformational sampling of the L9- $\alpha 4$  fragment. The bottom 4-hydroxyphenyl group connected to the CEI acetamide moiety is deeply buried in the active site cleft, where it is anchored in the slot tunnel through multiple hydrophobic (P224, I223 and I266) and aromatic  $\pi$ -stacking (W156) interactions. The 4-hydroxyphenyl group interacts with the indole NH group of W156 through a water-mediated hydrogen bond bridge. The acetamide moiety of CEI is positioned close to the conserved catalytic centre. In chain B, the top 4-hydroxyphenyl group linked with the CEI pyrazine ring interacts with a side chain of K189 through a water-mediated hydrogen bond bridge and forms a hydrogen bond with the carboxylate group of D162 from a symmetry-related enzyme molecule.

corresponding to the L9- $\alpha 4$  fragment from the modern RLuc8 enzyme into the ancestral Anc<sup>HLD</sup>-RLuc. The resulting AncFT protein had lower thermal stability than the ancestral variant (with ca. 7 °C lower  $T_m$  value) but folded correctly (with a CD spectrum similar to that of Anc<sup>HLD</sup>-RLuc and AncINS; Supplementary Fig. 13). We also constructed a second variant with L14 transplanted from RLuc8 onto the ancestral scaffold, but this protein easily aggregated, preventing its biochemical characterization.

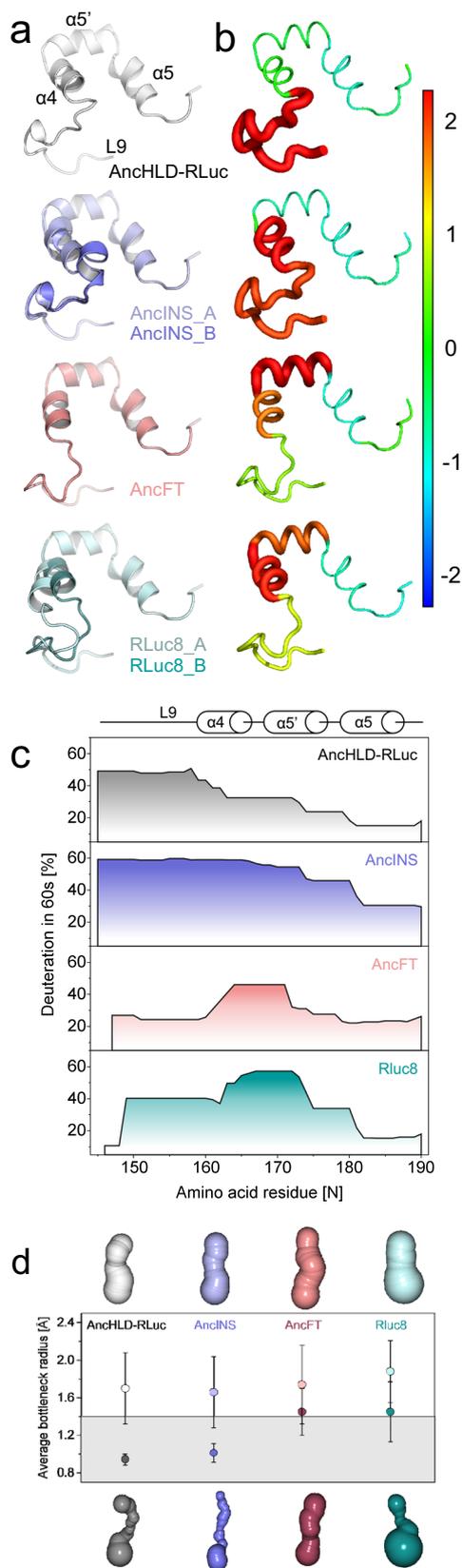
Transplantation of the L9- $\alpha 4$  flexible region from RLuc8 into the scaffold of Anc<sup>HLD</sup>-RLuc affected substrate-binding kinetics (Fig. 3): AncFT was found to have comparable initial collision kinetics to AncINS and RLuc8, but the fragment transplantation enhanced the following conformational change and (unlike AncINS) AncFT reached the overall binding efficiency of the modern luciferase. These findings correspond well with the results of the steady-state analysis. Similar values of the specificity constant  $k_{cat}/K_m$  were obtained for AncFT and RLuc8, indicating comparable efficiency of substrate binding. Thus, AncFT's lower LUC activity is due to a lower rate of the chemical transformation (i.e., the step after binding), as indicated by a 10-fold lower  $k_{cat}$  value for AncFT compared to RLuc8 (Fig. 3a, Supplementary Table 4). On the other hand, AncFT exhibited markedly weaker product inhibition than RLuc8. This was apparent from a significantly lower  $K_m/K_p$  ratio, quantifying a binding of the substrate and the product, which has an important effect on the bioluminescence signal's stability. The significant product inhibition of RLuc8, indicated by the high  $K_m/K_p$ , is one of the reasons why the bioluminescence signal provided by the luciferase rapidly decays after a strong initial flash. The fast inactivation and signal instability are major limitations of this popular molecular probe. Re-engineering of RLuc8 by ancestral reconstruction and subsequent backbone modification changed the unstable flash-type of bioluminescence to a significantly more stable glow-type bioluminescence. The half-life ( $t_{1/2}$ ) of AncFT bioluminescence was two orders of magnitude longer than that of RLuc8 (Supplementary Table 7, Fig. 7a, Supplementary Fig. 14), being consistent with a significantly lower  $K_m/K_p$  ratio determined for

AncFT, in comparison to RLuc8. The highly stable glow-type bioluminescence of AncFT persisted when this protein was expressed heterologously in mammalian (mouse fibroblast) cells (Fig. 7b, Supplementary Fig. 14). This result indicates that AncFT could be used as a reporter protein<sup>22</sup> for in vivo experiments.

Crystallographic analysis of AncFT (PDB ID 6S97, Supplementary Table 6) revealed a canonical  $\alpha/\beta$ -hydrolase fold similar to that of Anc<sup>HLD</sup>-RLuc, but with some structural features reminiscent of RLuc8. Most importantly, both the active site cavity and the  $\alpha 4$  helix conformation of AncFT are more open than in Anc<sup>HLD</sup>-RLuc, AncINS, and even the RLuc8 chain B (Fig. 4). Chain A of RLuc8 (PDB ID 2PSF) has the most open conformation seen in any of the structures (Figs. 4, 6a).

HDX-MS experiments showed that the deuteration pattern of AncFT was almost identical to that of Anc<sup>HLD</sup>-RLuc, with differences only in the transplanted region (Fig. 6c, Supplementary Fig. 11). Specifically, the end of the L9 loop was less extensively deuterated than in Anc<sup>HLD</sup>-RLuc while the  $\alpha 4$ - $\alpha 5'$  helices were more heavily deuterated. A similar trend was observed for RLuc8, providing experimental evidence that cap domain dynamics of the mutant with the transplanted fragment mimic those of RLuc8 (Fig. 6c).

MD simulations of AncFT revealed that the B-factor profile of its cap domain closely resembled that of RLuc8 (Fig. 6b). The B-factors were highest in the  $\alpha 5'$  and  $\alpha 4$  helices, while those of the L9 and L14 loops were close to the average for the rest of the enzyme and the rest of the cap domain had lower values (Supplementary Fig. 12). Tunnel analysis using Caver<sup>23</sup> showed that tunnel dynamics of Anc<sup>HLD</sup>-RLuc and AncINS clearly differ from those in AncFT and RLuc8 (Fig. 6d). The former two enzymes can adopt conformations with closed tunnels (average bottleneck radii: 0.9 and 1.0 Å, respectively), and conformations with open tunnels (average bottleneck radii: 1.7 Å in both cases). We consider a tunnel to be closed when its bottleneck radius drops below the radius of a water molecule (1.4 Å). In contrast, there is a clear overlap in bottleneck values calculated for the extremes of the conformational spectra of AncFT and RLuc8: the conformations that form the narrowest tunnels have identical



**Fig. 6 Engineering conformational dynamics by backbone modifications. a**

Conformational dynamics in the cap domains of crystal structures: AncINS (PDB ID 6S6E) has two monomers in the asymmetric unit differing in conformation of the  $\alpha 4$  helix carrying the insertion. Chains A (light blue) and B (slate) are in the open and closed conformations, respectively. RLuc8 (PDB ID 2PSF) has analogous conformations, with an open chain A (pale cyan) and a closed chain B (light teal). AncFT (PDB ID 6S97) has one monomer in the asymmetric unit, in which helix  $\alpha 4$  is more open than in Anc<sup>HLD</sup>-RLuc, AncINS, and even RLuc8 chain B. **b** Conformational dynamics in the cap domains observed during molecular dynamics simulations. B-factors of backbone atoms standardized across all protein variants, ranging from  $-2$  (blue) to  $2$  (red). B-factor values standardized for each protein are indicated by the thickness of the lines representing the protein backbone. Values were averaged per secondary structure element. **c** Dynamics and hydration of the cap domain based on HDX-MS assessments of peptide deuteration after 60 s. Anc<sup>HLD</sup>-RLuc is most heavily deuterated in the L9 loop. The deuteration pattern of AncFT is most similar to that of RLuc8. **d** Main access tunnel geometries of the proteins: representative snapshots of the open (lighter shades, upper row) and closed (darker shades, lower row) conformations. Average bottleneck radii and standard deviations from  $n = 1000$  snapshots from MD simulations were calculated by Caver 3.02<sup>23</sup>. The grey area corresponds to closed conformational states with tunnel radii below  $1.4 \text{ \AA}$  (the radius of a water molecule). Source data is available as a Source data file for Fig. 6b and c.

analysing 100 random snapshots from MD simulations we noticed that cavity volumes of all proteins can adopt wide ranges, but they also converge to similar minimal volume of  $200 \text{ \AA}^3$  in the closed state. The volumes calculated from the simulations starting from the open and closed conformations showed different absolute values, except for AncFT, which does not have multiple monomers in the asymmetric unit (Supplementary Fig. 15). Neither analysis of volumes of the static structures nor 100 representative snapshots obtained from MD simulations show any sign of correspondence between the catalytic characteristics and the active site volumes. Furthermore, on interpreting whether the system dynamics or the cavity volume is more important for the catalytic activity, the analyses of the ANM presented above agree with the importance of the dynamics.

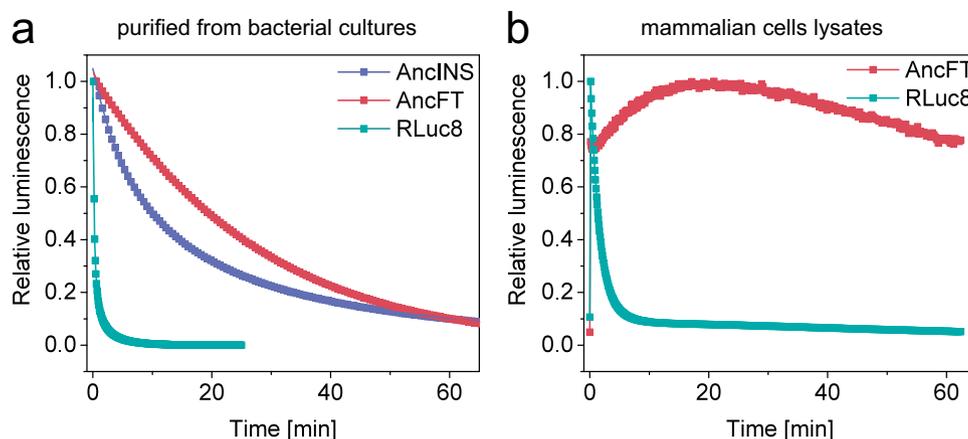
## Discussion

The present work validates a strategy for engineering enzyme backbone dynamics based on InDel mutagenesis of a stable (high melting temperature) and catalytically promiscuous (evolvable) template. We used TRIAD<sup>12</sup> mutagenesis to generate single amino acid insertion and deletion libraries of a stabilized bi-functional ancestral enzyme<sup>14</sup> of haloalkane dehalogenase and *Renilla*-type luciferase enzymes (Fig. 1). Alternatively, other published methods for introducing insertions and deletions could be used<sup>24–27</sup>. The most potent insertion variant, AncINS, has both an insertion and a substitution in the  $\alpha 4$  helix. These mutations markedly improved substrate binding, linked to changes in dynamics of the cap domain, but not the putative catalytic pentad. Remarkably, this was achieved in a single round of mutagenesis. In contrast, routinely used substitution mutagenesis protocols often require multiple time-consuming rounds of directed evolution and high-throughput screenings<sup>28–30</sup>. Systematic mutagenesis strategies, that allow a single-round selection to identify regions of interest for more focused study, are available<sup>31</sup>. In the future, maximizing information obtained from the selection through deep sequencing of gene libraries<sup>32</sup> may also provide information to predict beneficial mutations based on a single round.

The crystal structure of the insertion variant AncINS (PDB ID 6S6E) revealed that its asymmetric unit contains two monomers

average bottleneck radii of  $1.5 \text{ \AA}$ , while those forming open tunnels have average bottleneck radii of  $1.7$  and  $1.9 \text{ \AA}$ , respectively (Fig. 6d).

In addition, we performed an analysis of the cavity size of the proteins to test whether changes in the catalytic properties correlate with the changes in the volume of the active site. When



**Fig. 7 Comparison of bioluminescence of enzymes purified from bacterial cultures and in lysates from mammalian cells.** **a** Full decay kinetics of the conversion of 2.2  $\mu\text{M}$  CTZ by 50 nM of RLuc8, AncINS and AncFT purified from bacterial cultures. The data are presented as relative values to initial luminescence; Anc<sup>HLD-RLuc</sup> is not plotted due to low activity leading to large signal scattering. Solid lines represent the best fit to the experimental data. Experiments were repeated independently three times with consistent results. **b** Bioluminescence signal steadiness in relative values in lysates from mammalian cells expressing AncFT and RLuc8. Experiments were repeated independently three times with consistent results. Activity was measured using the commercial *Renilla* luciferase assay kit (Promega) with 20  $\mu\text{L}$  of cell lysates and 100  $\mu\text{L}$  of assay buffer. Luminescence signal in lysates from mammalian cells expressing Anc<sup>HLD-RLuc</sup> and AncINS was not detectable under tested conditions. Source data is available as a Source data file for Fig. 7.

with different spatial arrangements of the  $\alpha 4$  helix (implying they represent two conformers), and MD simulations confirmed that the mutations caused substantial structural rearrangements, most probably because the substituted residue P163 acts as a helix breaker. We conclude that the LUC activity benefits from conformational dynamics of the protein's rigid scaffold to allow binding of a bulky substrate. The outward movement of the  $\alpha 4$  helix has an immediate effect on the opening of the access tunnel and increasing of the buried cavity's volume, thus facilitating binding of the bulky substrate CTZ. However, once CTZ is bound inside of the active site cavity, the protein has to provide tight binding of the molecule inside it and provide a desolvated environment for the subsequent monooxygenation reaction and bioluminescence to occur<sup>33,34</sup>. If the excited product of the reaction, CEI, would not be tightly bound, the energy would probably not be released in the form of light during decay to the ground state, but in another form, e.g., heat. Accordingly, we propose that the evolutionary processes that generated the modern *Renilla* luciferase altered its dynamics in a way that facilitated substrate binding to the open state. While the movement of the  $\alpha 4$  helix was shown to be important for CTZ binding, the dynamic profile of AncINS, as determined by HDX-MX experiments and MD simulations, still differs from that of RLuc8. To validate the proposed conclusions, we transplanted the  $\alpha 4$  helix and L9 loop from the highly efficient modern RLuc8 into the poorly active ancestral Anc<sup>HLD-RLuc</sup>. The resulting enzyme, AncFT, was highly efficient, with significantly decreased product inhibition compared to RLuc8, providing glow-type luminescence. Crystallographic analysis revealed that its structure (PDB ID 6S97) was more open than those of both the ancestral protein and B chain of RLuc8, facilitating the entrance of the bulky substrate. Based on MD simulations and HDX-MS experiments, we conclude that the transplantation reduced motions of the L9 loop, but increased motion of the  $\alpha 4$ - $\alpha 5'$  helices, making the mutant protein's dynamics more similar to RLuc8.

Our results provide experimental evidence for the role of protein dynamics in enzymatic catalysis. An ancestral protein (Anc<sup>HLD-RLuc</sup>) with sub-optimal dynamics has its catalytic activity limited by substrate binding, while mutants with tailored dynamism resulting from insertion/substitution (AncINS) or

fragment transplantation (AncFT) have catalytic activity limited by the subsequent step of the chemical reaction. Transplantation of the most dynamic region of the extant RLuc onto the ancestral Anc<sup>HLD-RLuc</sup> improved its catalytic efficiency 7000-fold to match the catalytic efficiency of RLuc, while extending the half-life of its light output 100-fold. Long-lasting glow-type light emission of AncFT was confirmed in mammalian cells, which paves the way towards its use as an efficient molecular reporter or biosensor<sup>22</sup>. Given the growing sensitivity of optical devices, signal stability is becoming a key requirement for modern molecular probes. In comparison with the short-term flash type of signal, a stable glow signal enables continuous detection in long-term biological experiments or maintenance of a stable response required during high-throughput screening campaigns<sup>22</sup>.

Thus, both the introduction of InDels and transplantation of dynamic elements into stable ancestors appear to be viable protein engineering strategies for improving enzymes or introducing novel functions. The potential effectiveness of grafting dynamic loops to introduce novel enzymatic functions has been discussed<sup>35–37</sup> and experimentally demonstrated by the successful introduction of  $\beta$ -lactamase activity into the  $\alpha\beta/\beta\alpha$  metallohydrolase scaffold<sup>38</sup>, thereby generating an enzyme that lacked its original activity but catalysed hydrolysis of cefotaxime. The potential of loop remodelling for engineering enzyme functions has also been illustrated by the deletion of a specific loop and introduction of a point mutation, leading to emergence of homoserine lactonase activity in a phosphotriesterase<sup>39</sup>. The potential of loop modification in enzyme engineering has also been discussed in several recent reviews<sup>35–37</sup>.

There are several preconditions necessary for the application of our engineering strategy. Ancestral sequence reconstruction requires about 150 sequences<sup>40</sup>, but a phylogenetic tree can also be reconstructed with a lower number of sequences. Mutational events leading to insertions and deletions make the reconstruction less reliable<sup>41</sup>. The InDel mutagenesis strategy is applicable to any protein of interest, but a screening assay is necessary for probing a sufficient number of mutants to identify the regions of interest and to allow robust statistical analysis. For the loop transplantation strategy, structural information about the proteins is necessary. The strategy is applicable for engineering

catalytic efficiency, substrate specificity, and selectivity, and may enable the design and evolution of multifunctional enzymes. It is specifically suitable for enzymes with active sites flanked by loops and enzymes with reaction rates limited by substrate binding or product release.

In summary, we have developed a strategy for engineering backbone dynamics based on InDel mutagenesis of a stable template, multivariate statistics of the data from microscale and microfluidic experiments, and transplantation of a highly dynamic element. We validated its utility by an engineering effort in which catalytic efficiency was increased, and the bioluminescence was stabilized to achieve a long-lasting glow-type signal. The results achieved support the conceptual ideas, which guided our experiments and stepwise implementation of the strategy. The developed catalyst can serve as a molecular probe in bacterial and mammalian cells. The strategy may provide a useful addition to the repertoire of methods available for engineering catalytically efficient enzymes<sup>40,41</sup> by exploring less travelled parts of protein sequence space. The loop transplantation strategy is being implemented in the web application LoopGrafter, which will make it accessible to a broad community (<https://loschmidt.chemi.muni.cz/loopgrafter/>).

## Methods

**Reagents and procedures used in TRIAD.** FastDigest restriction endonucleases, MuA transposase and T4 DNA ligase were purchased from Thermo Fisher Scientific. DNA Polymerase I, Large (Klenow) Fragment, was purchased from New England Biolabs. All DNA modifying enzymes were used according to the manufacturer's conditions. All DNA purification procedures were performed according to the manufacturers' instructions using kits, including GeneJET Plasmid Miniprep kit (Thermo Fischer Scientific) for plasmid extractions from *E. coli* cells, Zymo-clean Gel DNA Recovery kit (Zymo Research) for agarose gel extraction of DNA fragments upon electrophoresis and DNA Clean & Concentrator kit (Zymo Research) for DNA purification and concentration. Bacterial transformations were performed by electroporation using *E. coli* 10G ELITE electrocompetent cells (Lucigen).

### Generation of insertion and deletion libraries of Anc<sup>HLD-RLuc</sup> using TRIAD.

InDel libraries were prepared following the TRIAD<sup>12,42</sup> method. The sequence encoding Anc<sup>HLD-RLuc</sup> was subcloned from pET21b::anc<sup>HLD-RLuc</sup> to the TRIAD-dedicated vector pID-Tet using NdeI and BamHI and yielding pID-Tet::anc<sup>HLD-RLuc</sup>. This construct was used as template for the generation of transposition insertion libraries with engineered TRIAD transposons TransDel and TransIns. The transposons (~1 kbp) were extracted from pUC57 by BglII digestion and recovered by gel electrophoresis and purification. Insertion of TransDel or TransIns in pID-Tet::anc<sup>HLD-RLuc</sup> (pID-Tet plasmid: ~2.7 kbp; anc<sup>HLD-RLuc</sup>: ~950 bp) was performed by in vitro transposition using ~300 ng of plasmid, ~50 ng of transposon and 0.22 µg MuA transposase in a 20 µL reaction volume. After incubation for 2 h at 30 °C, the MuA transposase was heat-inactivated for 10 min at 75 °C. DNA products were purified and concentrated in 7 µL deionized water. Two microlitres of the purified DNA was used to transform *E. coli* 10G ELITE electrocompetent cells by electroporation. The transformants (~30,000–50,000 CFU) were selected on LB agar containing ampicillin (amp; 100 µg/mL) and chloramphenicol (cam; 34 µg/mL). The resulting colonies were pooled, and their plasmid DNA extracted. The fragments corresponding to anc<sup>HLD-RLuc</sup> containing the inserted transposon (~2 kbp) were obtained by double restriction digestion (NdeI/BamHI) followed by gel extraction and ligated back into pID-Tet (50–100 ng). The ligation products were then transformed into *E. coli* 10G cells. Upon selection on LB-agar-amp-cam, transformants (~1–2 × 10<sup>6</sup> CFU) were pooled and their plasmid DNA extracted, yielding TransDel or TransIns insertion libraries depending on the TRIAD transposon used at the start. For the generation of the triplet nucleotide deletion library of Anc<sup>HLD-RLuc</sup>, TransDel insertion library plasmids were first digested with MlyI to remove TransDel. The fragments corresponding to linearized pID-Tet::anc<sup>HLD-RLuc</sup> (with a ~3 bp deletion in Anc<sup>HLD-RLuc</sup>) were isolated by gel electrophoresis and purified. Self-circularization was then performed using T4 DNA ligase and 10–50 ng linearized plasmid in 50 µL reaction volume (final DNA concentration: ≥1 ng/µL). Upon purification and concentration, the ligation products were transformed into electrocompetent into *E. coli* 10G cells subsequently selected on LB-agar-amp, yielding a library of gene of interest variants with random triplet nucleotide deletions. For the generation of the triplet nucleotide insertion library of Anc<sup>HLD-RLuc</sup>, TransIns insertion library plasmids were first digested with NotI and MlyI to remove TransIns. The linearized pID-Tet::anc<sup>HLD-RLuc</sup> plasmids were recovered by gel electrophoresis and purification. TRIAD cassette Ins1 (containing randomized 3 bp at one extremity and a kanamycin-resistance gene) was extracted from pUC57 by

NotI/MlyI digestion, isolated by gel electrophoresis, purified and inserted into the linearized pID-Tet::anc<sup>HLD-RLuc</sup> plasmid (50–100 ng) in a 1:3 molar ratio. After purification and concentration, these ligation products were transformed into *E. coli* 10G cells and the transformants (~2 × 10<sup>6</sup> CFU) were selected on LB-agar supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin. The resulting plasmid ins1 insertion library was then extracted from the transforming colonies and subsequently digested with AclI. The linearized pID-Tet::anc<sup>HLD-RLuc</sup> plasmids (with a 3 bp insertion) were recovered by gel electrophoresis, purified and subsequently treated with the Klenow fragment of DNA Polymerase I to remove 3' overhangs created by AclI digestion. After that blunting step, the plasmids were self-circularized. The resulting ligation products were transformed into electrocompetent *E. coli* 10G cells subsequently plated on LB-agar-amp, yielding libraries of Anc<sup>HLD-RLuc</sup> variants with random triplet nucleotide insertions (one per variant). The same procedure was applied to generate a second-round triplet nucleotide insertion library of AncINS. All the libraries were purified and stored in the form of plasmid solutions prior to transformation into *E. coli* for protein variant expression and screening. Plasmid DNA from 10 randomly selected colonies from all three libraries were sequenced for quality control (Eurofins Genomics, Germany).

**Point mutagenesis of RLuc8.** Site-directed PCR-based mutagenesis was applied to create *rLuc8-W121F/E144Q* in two steps using QuikChange site-directed mutagenesis kit via manufacturer's protocol (Agilent, USA). First, *rLuc8-W121F* was created using oligonucleotides RLuc8-W121F-FWD1 and RLuc8-W121F-RVS1 with *rLuc8* gene serving as a template. Then *rLuc8-W121F* gene was used as a template to create *rLuc8-W121F/E144Q* using RLuc8-E144Q-FWD2 and RLuc8-E144Q-RVS2. Error-free clones were confirmed by DNA sequencing (Eurofins Genomics, Germany). Mutagenic primers are available in Supplementary Table 8.

**Transplantation of secondary structure elements.** The Anc<sup>HLD-RLuc</sup> and RLuc8 sequences were aligned using Jalview<sup>43</sup> and T-coffee<sup>44</sup> with default settings (Supplementary Fig. 9). The construct pET21b::anc<sup>HLD-RLuc</sup> (NdeI/BamHI) was chosen as a template for mutagenesis using Phusion® High-Fidelity DNA Polymerase according to the manufacturer's protocol (New England Biolabs, USA). Mutagenic primers (Supplementary Table 8) (Sigma-Aldrich, USA) were designed manually for two separate PCR runs (Supplementary Fig. 16). In the next step, a standard fusion PCR protocol was used to fuse DNA fragments from the first two PCR runs. The resulting fused DNA fragment was cloned into pET21b or pID-Tet vectors. The error-free status of the clones was confirmed by sequencing (Eurofins Genomics, Germany).

**Cultivation of InDel libraries in 96-well plates.** Chemocompetent *Escherichia coli* BL21 cells (#C2530H, New England BioLabs, USA) were transformed with the anc<sup>HLD-RLuc</sup> insertion/deletion library (in the pID-Tet vector) and grown on LB agar plates containing 100 µg/ml ampicillin overnight at 37 °C. The cells transformed with the vector pID-Tet (lacking an insert) and pID-Tet::anc<sup>HLD-RLuc</sup> were used as negative and positive controls, respectively, when screening the libraries. Single colonies of transformed cells carrying InDel variants of anc<sup>HLD-RLuc</sup> were transferred into sterile 96-well plates containing 150 µL LB medium with 100 µg/ml ampicillin in each well. The plates were covered with AeraSealTM film (Sigma-Aldrich, USA) and incubated for 15 h at 37 °C under shaking at 200 rpm. After cultivation, 100 µL of the culture was transferred into new microtiter plates (MTPs) and 100 µL of fresh LB medium with 100 µg/ml ampicillin and anhydrotetracycline (to a final concentration 200 ng/ml) was added to each well. To make replica plates to be stored at -70 °C, 50 µL of 30% glycerol was added to the remaining 50 µL of the culture. The MTP was incubated at 30 °C and 200 rpm for 4 h. Cell cultures were harvested by centrifugation at 1600 × g for 20 min. The supernatant was discarded and the MTP was frozen at -70 °C. Before screening of the libraries, MTPs were defrosted and kept at laboratory temperature for 10 min. Then, 70 µL of lysis buffer (20 mM potassium phosphate, 20 mM Na<sub>2</sub>SO<sub>4</sub> and 1 mM EDTA, pH 8.0) containing lysozyme (1 mg/ml) was added to each well. Cell debris was removed from the lysate by centrifugation at 1600 × g for 20 min after incubation at 23 °C and 100 rpm for 1 h. Robotic cultivation and cell lysis were performed analogously to the manual screening described above using a colony picking robot (Colony Picker, Molecular Devices QPix 420), a liquid handling robot (Bravo, Agilent) with a 96-tip head and the LARA robotic system (Greifswald, Germany).

**Screening of luciferase activity in 96-well plates.** Luciferase activity was measured using a previously described procedure<sup>14</sup>, with adaptation, in both the manual and robotic screenings. For manual screening, 30 µL of cell lysate was transferred into a new MTP and 220 µL of assay buffer (100 mM potassium phosphate, 1 mM Na<sub>2</sub>SO<sub>4</sub>, pH 7.5) with 2.2 µM coelenterazine (CTZ) was added. The luminescence signal was immediately measured for 22 s with the gain value set to 3250. Before the addition of the assay buffer with CTZ, the sample's baseline luminescence signal was measured for 10 s. Luminescence was measured at 30 °C in a FLUOstar OPTIMA Microplate Reader (BMG Labtech, Germany). Luciferase activity was expressed in relative light units (RLU) s<sup>-1</sup> mg<sup>-1</sup> of an enzyme<sup>14</sup>. The RLU were integrated over the first 72.5 s immediately after injection of the substrate into the enzyme solution.

For robotic screening, 50  $\mu$ l of cell lysate was pipetted by a Liquid Handling Robot into a new white MTP (Nunc™ F96 MicroWell™ White Polystyrene Plate, Nunclon Delta Surface, Thermo Fisher Scientific) and the reaction was initiated by adding 50  $\mu$ l of a CTZ stock solution (1 mg of CTZ dissolved in 2 ml of pure EtOH) diluted 1000x in MilliQ water. The luminescence signal was measured using a VarioscanLUX reader (Thermo Fisher Scientific, SkanIt Software 4.1 for Microplate Readers RE, ver. 4.1.0.43). Each measurement consisted of 50 readings, each of which took 200 ms (standard optics, automatic dynamic range). The CTZ solution was dispensed during reading 15 at a moderate dispensation speed. Luciferase activity was expressed in relative light units (RLU)  $s^{-1} mg^{-1}$  of an enzyme<sup>14</sup>.

**Screening of dehalogenase activity in 96-well plates.** Haloalkane dehalogenase activity measurements were based on a pH assay<sup>45</sup> according to Holloway and co-workers<sup>45</sup>. The principle of the assay is based on the detection of protons produced during the dehalogenation reaction. Substrate 1-bromobutane (100  $\mu$ l) was incubated in the reaction buffer (200 ml; 1 mM HEPES, 20 mM  $Na_2SO_4$ , 1 mM EDTA and 25  $\mu$ g/ml phenol red, pH 8.2) at 37 °C for 30 min. Fifteen microlitres of cell lysate was transferred into new MTP and 185  $\mu$ l of assay buffer with 1-bromobutane were added. The MTP plate was properly closed by the Adhesive film for microplates (VWR, USA) and incubated at laboratory temperature for 15 h. The change in colour of pH indicator was measured at 540 nm by using an Eon spectrometer (BioTek, USA).

**Overproduction and purification of Anc<sup>HLD</sup>-RLuc variants.** Anc<sup>HLD</sup>-RLuc variants were overexpressed from pID-Tet (amp<sup>R</sup>) in *E. coli* BL21 cells (#C2530H, New England BioLabs, USA) cultivated in LB medium supplemented with ampicillin (100  $\mu$ g/ml) at 37 °C. Protein production was induced under the TET promoter at 20 °C once the OD<sub>600</sub> reached ~0.5 by adding anhydrotetracycline (Cayman Chemical, USA) to a final concentration of 200 ng/ml. At small scale, proteins were purified using the MagneHis™ Protein Purification System (Promega, USA), dialysed using a Slide-A-Lyzer™ MINI Dialysis Device (Thermo Scientific, USA), and their purity was verified by SDS-polyacrylamide gel electrophoresis. At large scale, proteins were purified by affinity chromatography targeting their C-terminal hexahistidine tags. The monomer fraction was separated on a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare, UK) equilibrated with 100 mM potassium phosphate buffer (pH 7.5). All enzymes were concentrated using Amicon<sup>R</sup> Ultra-15 Ultracel<sup>R</sup>-10K Centrifugal Filter Units (Merck Millipore Ltd., Ireland). The purity of all enzyme preparations was checked by SDS-polyacrylamide gel electrophoresis; in all cases, only one band corresponding to the monomer fraction was visible.

**Large-scale overproduction and purification of RLuc8-W121F/E144Q.** The *E. coli* BL21 (DE3) (C2527H, New England BioLabs, USA) cells were transformed by the heat shock method with the plasmid pET21b::rLuc8-W121F/E144Q. Ampicillin-resistant colonies were inoculated in LB medium (10 ml) (1xLB medium, ampicillin 100  $\mu$ g/ml), which was incubated (200 rpm) overnight at 37 °C. On the next day, a bacterial culture was used to inoculate 5-liter Erlenmeyer flasks containing 1 liter of 1xLB medium with ampicillin (100  $\mu$ g/ml) where cells were grown (200 rpm, 37 °C) until the culture reached OD<sub>600</sub> = 0.5. Induction of expression was done at 22 °C by adding 0.5 ml of 1 M IPTG, and the culture was then incubated overnight (typically 12–16 h) at 22 °C/115 rpm. Next day, the bacterial biomass was harvested by centrifugation (3000  $\times$  g/10 min/4 °C) and re-suspended in a purification buffer (500 mM NaCl, 20 mM potassium phosphate buffer pH 7.5, 10 mM imidazole) and lysed by sonication (50% amplitude, 32 min (5 s pulse/5 s pause) using a sonicator Sonic Dismembrator Model 705 (Fisher Scientific, USA). The sonicated lysate was clarified by centrifugation (21,000  $\times$  g, 60 min, 4 °C). The supernatant was then collected, filtered and applied on Ni-NTA Superflow Cartridge (Qiagen, Germany) column equilibrated with the purification buffer. The target enzyme was eluted by a linear gradient of the purification buffer supplemented with 300 mM imidazole. The eluted protein fractions of RLuc8-W121F/E144Q were pooled and dialysed overnight against 100 mM NaCl, 10 mM Tris-HCl pH = 7.0. The dialysed protein was subsequently loaded onto 16/60 Superdex 200 gel filtration column (GE Healthcare, UK) pre-equilibrated with the corresponding buffer. The purified RLuc8-W121F/E144Q protein was concentrated with an Amicon Ultra centrifugal filter units (Millipore, USA), and protein concentrations were assayed by the DS-11 Spectrophotometer (DeNovix, USA).

**Mammalian cells experiments.** NIH/3T3 mouse fibroblast cells (ATCC<sup>®</sup> CRL-1658™) were transfected according to manufacturer's protocol using Lipofectamine 2000 (Thermo Fisher, USA) with pcDNA3.1(+) plasmids containing genes codon-optimized for expression in mammalian cells (Gene Art, Thermo Fisher, USA). Cells were lysed 24 h after transfection and luciferase activity was measured in lysate via Microplate Reader FLUOstar Omega (BMG Labtech, Germany) using a commercial Renilla Luciferase Assay System (Promega, USA) and also using an in-house prepared assay buffer 100 mM PBS pH = 7.5 with 4.5  $\mu$ M CTZ (final concentration in the reaction mixture). Cells transfected with pcDNA3.1(+) plasmid were used as a negative control. The measurements were done in triplicates.

**Circular dichroism spectroscopy.** Circular dichroism (CD) spectra were recorded at 20 °C using a Chirascan spectropolarimeter (Applied Photophysics, UK). Data were collected from scans of the focal proteins from 185 to 260 nm at 100 nm/min with a 1 s response time and 1 nm bandwidth in 0.1 cm quartz cuvettes. Each presented spectrum (Fig. S13) is an average of five individual scans, corrected for absorbance of the buffer. The CD data were expressed in terms of mean residue ellipticity  $\Theta_{MRE}$  described by Eq. (1).

$$\Theta_{MRE} = \frac{\Theta_{obs} \cdot M_w \cdot 100}{n \cdot c \cdot l} \quad (1)$$

where  $\Theta_{obs}$  is the observed ellipticity in degrees,  $M_w$  is the scanned protein's molecular weight,  $n$  is the number of residues,  $l$  is the cell path length,  $c$  is the protein concentration (in mg/ml) and the factor of 100 originates from the conversion of the molecular weight to mg/dmol.

**Thermal stability.** Thermal unfolding was studied by using a NanoDSF Prometheus instrument (NanoTemper, Germany) to monitor Trp fluorescence during heating at 1 °C/min from 20 to 90 °C. The melting temperatures ( $T_{onset}$  and  $T_m$ ) were evaluated directly by ThermControl v2.0.2.

#### Microfluidic determination of temperature profiles and thermodynamics.

Temperature profiles of specific activities of individual enzyme variants towards the substrate 1,3-dibromopropane were measured in 2.5 °C increments from 20 to 40 °C using a capillary-based droplet microfluidic platform, enabling characterization of enzymatic activity within droplets for multiple enzymes in a single run<sup>46</sup>. The droplets were generated using the Mitos Dropix (Dolomite, UK). A custom sequence of droplets (150 nl aqueous phase, 300 nl oil spacing) was generated using negative pressure (microfluidic pump) and the droplets were guided through polyethylene tubing to the incubation chamber. Within the incubation chamber, the halogenated substrate was delivered to the droplets via a combination of microdialysis and partitioning between the oil (FC 40) and the aqueous phase. The reaction solution consisted of a weak buffer (1 mM HEPES, 20 mM  $Na_2SO_4$ , pH 8.2) and a complementary fluorescent indicator 8-hydroxypyrene-1,3,6-trisulfonic acid (50  $\mu$ M HPTS). The fluorescence signal was obtained by using an optical setup with excitation laser (450 nm), a dichroic mirror with a cut-off at 490 nm filtering the excitation light and a Si-detector. By employing a pH-based fluorescence assay, small changes in the pH were observed and enabling monitoring of the enzymatic activity. The reaction progress was analysed as an end-point measurement recorded after passing of 7 or 10 droplets/sample through the incubation chamber. The reaction time was 4 min. The raw signal was processed by a droplet detection script<sup>46</sup> written in MATLAB 2017b (Mathworks, USA) to obtain the specific activities. Natural logarithms of the specific activities were used in subsequent thermodynamic analysis to generate both Arrhenius and Eyring plots and to derive thermodynamic parameters (Supplementary Fig. 1).

**Anisotropic network modelling.** Secondary structure elements were defined in Anc<sup>HLD</sup>-RLuc and RLuc8 based on their respective crystal structures (PDB ID 6G75, chain A; and 2PSF, chain B) using DSSP<sup>47</sup> followed by manual edition after visual inspection (Supplementary Tables 2, 3). Anisotropic network models were computed using the Prody 1.10.8 standalone package<sup>48</sup> and the position-specific vector of squared fluctuations and matrix of motion cross-correlations were obtained. These matrices were further extended by calculating the averaged cross-correlation values corresponding to each secondary structure element. To facilitate distinction of differences in predicted motions, the matrix calculated for RLuc8 was subtracted from that calculated for Anc<sup>HLD</sup>-RLuc. The final matrix  $M$  encompassed values ranging from -2 (motions more cross-correlated in RLuc8) to 2 (motions more cross-correlated in Anc<sup>HLD</sup>-RLuc), where values close to 0 indicate similar motions in the two structures. To improve understanding of effects of mutations on the 25 selected variants, the positions and secondary structure elements of mutations resulting in extreme values were mapped to both reference crystals (Anc<sup>HLD</sup>-RLuc and RLuc8) based on their alignment (Supplementary Fig. 9). Then, three groups of interesting regions were defined and used to annotate each variant with 18 values obtained from matrix  $M$ : nine for the cross-correlation of motions of the specific position and nine for the secondary structure element.

**Calculation of segment-centred cross-correlation values from ANM.** Let  $i_1$  and  $i_2$  be the indices of the beginning and ending position of a given secondary structure element (SSE <sub>$i$</sub> ) in the structure for which the ANM is computed;  $j_1$  and  $j_2$  be the indices of the beginning and ending position of a different secondary structure element (SSE <sub>$j$</sub> ) on the same structure;  $var(x)$  and  $var(y)$  be the squared fluctuation value for the  $x$ th and  $y$ th residue on the structure, respectively; and  $r(x, y)$  the cross-correlation value for the pair formed by the  $x$ th and  $y$ th residues on the structure obtained from the same ANM calculation. The averaged cross-correlation

value for the pair ( $SSE_i$ ,  $SSE_j$ ) can be calculated according to Eq. (2).

$$\text{Ccor}(SSE_i, SSE_j) = \frac{\sum_{x=1}^E \sum_{y=1}^P (r(x, y) \sqrt{\text{var}(x) \cdot \text{var}(y)})}{\sqrt{\sum_{x=1}^E \sum_{y=1}^P (r(x, y) \sqrt{\text{var}(x) \cdot \text{var}(y)}) \cdot \sum_{x=1}^E \sum_{y=1}^P (r(x, y) \sqrt{\text{var}(x) \cdot \text{var}(y)})}} \quad (2)$$

Note that if additionally to the  $E$   $SSE_x$  elements defined in each of the Supplementary Tables S2, S3, each of the  $P$  amino-acids of each analysed protein is considered as one of such elements, the dimension of the resulting cross-correlation matrix is  $(E + P)^2$ . Such matrix would be formed by two diagonal matrices with dimensions  $E^2$  and  $P^2$  representing the SSE to SSE and amino-acid to amino-acid cross-correlations, respectively; and by two mirroring rectangular matrices ( $E \times P$  and  $P \times E$ ) representing the amino-acid to SSE cross-correlations.

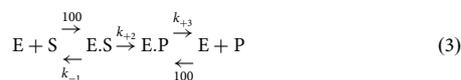
**Multivariate statistical analysis.** Partial least squares (PLS) regression analysis<sup>49</sup> was used to explore relationships between structural and molecular variables (X) of the proteins and their enzymatic activities (Y), all of which were autoscaled and centred. The complete data matrix used in the PLS analysis is provided in Source data file. Variable importance in the projection (VIP)<sup>50</sup> and variable weight plots<sup>51</sup> were used to assess the importance of every descriptor in the model, with further validation by cross-validation<sup>49</sup> and permutation testing. During the cross-validation procedure, parts of the Y data are not considered during model building, and the ability of the resulting model to predict those data is assessed by comparing predicted and actual values, providing cross-validated  $Q^2$  values. In this study, 1/7 of the mutants were deleted in each cross-validation round. In the permutation testing, the model was recalculated 999 times with a randomly re-ordered dependent variable. SIMCA-P version 12 (Umetrics, Umea, Sweden) was used for statistical analyses.

**Success rate frequency analysis.** Luminescence readings were processed to determine each tested variant's luciferase activity relative to the mutagenesis template. Each microtiter plate used for this analysis included negative and positive controls (in sets of four wells) as well as the tested variants, and transformations were applied to obtain relative activities of the preparation in each well. First, the negative control wells were averaged time-point by time-point. Second, the rest of the wells were blanked using subtracting from each of their time-point values of the corresponding averaged time-point from the negative controls. Third, two segments for each data series were defined: pre- and post-injection of the coelenterazine substrate, which occurred ten seconds after the beginning of the readings on the FLUOstar OPTIMA instrument. The pre-injection series consisted of all data points before injection time. The post-injection series consisted of all data points collected from the moment the reading of the signal was stable (0.2 s after the injection time) until the end of the readings. From each data series, outliers defined as  $Q1 - 1.5 \cdot (Q3 - Q1)$  and  $Q3 + 1.5 \cdot (Q3 - Q1)$  (where  $Q1$  and  $Q3$  represent the first and third quartile values of each distribution) were omitted to remove noise from the data. Fourth, the average value of the pre-injection series was subtracted from the average value of the post-injection one, to obtain the raw intensity value of each well. Fifth, the reference intensity value of the plate was calculated by averaging the raw intensity values of the positive control wells. Sixth, the relative intensity value was calculated as the ratio of the raw intensity value of each well over the reference intensity value of the plate.

**Steady-state kinetics measurement and data analysis.** Solid coelenterazine was dissolved in ice-cold ethanol and stored under nitrogen atmosphere in dark glass vials at  $-20^\circ\text{C}$ . Before measurement, concentration and quality of the ethanol stock solution were verified spectrophotometrically. Series of buffer solutions with different coelenterazine concentration was prepared by manual injection of an appropriate volume of the ethanol stock solution into 10 ml of 100 mM phosphate buffer pH 7.5 immediately before the measurement. The reaction mixture was composed of 10% (v/v) enzyme solution in 100 mM phosphate buffer pH 7.5 and 90% (v/v) of buffer solution of coelenterazine. All reactions were carried out at  $37^\circ\text{C}$  in microtiter plates using the microplate reader FLUOstar OPTIMA (BMG Labtech, Germany) set to broad-spectrum luminescence reading. Microplate well with pre-pipetted 25  $\mu\text{l}$  of enzyme solution was first monitored for background light for 10 s, after which, 225  $\mu\text{l}$  of a buffer solution with coelenterazine was added via an automatic syringe. The luminescence of the reaction mixture was then measured for the desired time until the luminescence intensity decreased under 0.5% of its maximal measured value. Each reaction was performed in three repetitions. The gain of the reader was tailored to each enzyme separately, however, for each enzyme, all readings at different substrate concentrations were obtained using the same gain value.

An updated protocol applying new standards for collecting and fitting steady-state kinetic data<sup>52,53</sup> was used. Unlike the classical initial velocity analysis, which requires a sophisticated luminometer calibration and quantum yield evaluation to obtain complete kinetic data<sup>54</sup>, luminescence data were recorded when letting the reaction proceed to completion beyond the initial linear phase. The recorded luminescence traces (rate vs. time) were transformed to reaction progress curves corresponding to cumulative luminescence in time (Supplementary Fig. 5). The transformed steady-state kinetic data (product vs. time) were fitted globally by

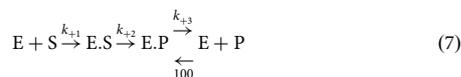
numerical methods with the KinTek Explorer (KinTek Corporation, USA) to obtain direct estimates of turnover number  $k_{\text{cat}}$ , Michaelis constant  $K_m$ , specificity constant  $k_{\text{cat}}/K_m$ , and equilibrium dissociation constant for enzyme-product complex  $K_p$  with no need for luminometer quantum yield calibration. The software allows for the input of a given kinetic model via a simple text description, and the programme then derives the differential equations needed for numerical integration automatically. Numerical integration of rate equations searching a set of kinetic parameters that produce a minimum  $\chi^2$  value was performed using the Bulirsch–Stoer algorithm with adaptive step size, and nonlinear regression to fit data was based on the Levenberg–Marquardt method<sup>55</sup>. To account for fluctuations in experimental data, enzyme or substrate concentrations were slightly adjusted ( $\pm 5\%$ ) to derive best fits. Residuals were normalized by sigma value for each data point. The standard error (S.E.) was calculated from the covariance matrix during nonlinear regression. In addition to S.E. values, more rigorous analysis of the variation of the kinetic parameters was accomplished by confidence contour analysis by using FitSpace Explorer (KinTek Corporation, USA). In these analyses, the lower and upper limits for each parameter were derived (Supplementary Table 4) from the confidence contour obtained from setting  $\chi^2$  threshold at 0.95<sup>56</sup>. The scaling factor, relating luminescence signal to product concentration, was applied as one of the fitted parameters, well constrained by end-point levels of kinetic traces recorded at particular substrate concentrations. Depletion of the available substrate after the reaction was ensured by repeated injection of the fresh enzyme. The steady-state model (Eqs. (3)–(6)) was used to obtain the values of turnover number  $k_{\text{cat}}$ , Michaelis constant  $K_m$  and equilibrium dissociation constant for enzyme-product complex  $K_p$ . A conservative estimate for diffusion-limited substrate and product binding  $k_{+1}$  and  $k_{-3}$  ( $100 \mu\text{M}^{-1}\cdot\text{s}^{-1}$ ) was used as a fixed value to mimic rapid equilibrium assumption. An alternative form of the model (Eqs. (7) and (8)) was used to obtain estimates of  $k_{\text{cat}}/K_m$  directly by setting  $k_{-1} = 0$ <sup>52,53</sup>.



$$k_{\text{cat}} = k_{+2} \quad (4)$$

$$K_m = k_{-1}/100 \quad (5)$$

$$K_p = k_{+3}/100 \quad (6)$$



$$k_{\text{cat}}/K_m = k_{+1} \quad (8)$$

**Bioluminescence decay kinetics.** Bioluminescence decay kinetics were monitored in the same conditions as the steady-state kinetics, with final concentrations of 2.2  $\mu\text{M}$  CTZ and 50 nM enzyme. The obtained luminescence kinetic data were fitted to an exponential decay or logistic model using Origin 6.1 (OriginLab, USA) to obtain the half-life  $t_{1/2}$  of the luminescence signal.

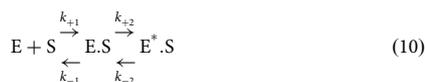
#### Transient kinetics of substrate binding experiments and data analysis.

Transient kinetic traces of the protein variants were monitored after rapidly mixing components in 100 mM potassium phosphate buffer (pH 7.5) using the Stopped-Flow SFM 3000 mixing system and the MOS-500 spectrometer (BioLogic, France). In each case, the reaction was initiated by mixing 75  $\mu\text{l}$  of an enzyme solution with 75  $\mu\text{l}$  of CTZ solution and then monitored by the changes of fluorescence at  $340 \pm 13 \text{ nm}$  after excitation at 295 nm. In experiments with each protein variant, concentrations of the CTZ substrate and an enzyme were separately varied in sets of seven consecutive replicates and the resulting kinetic traces were averaged. The experiments were based on monitoring changes in native tryptophan fluorescence that was quenched by the coelenterazine substrate upon its binding by an enzyme. The drop of the initial fluorescence level caused by the non-specific interactions and inner filter effects of coelenterazine was corrected based on experiments with bovine serum albumin, which exhibited the same initial decrease of the fluorescence signal. At room temperatures, the kinetics of the substrate binding was too fast, and most of the kinetic information was lost in the dead time of the instrument (0.3 ms). After decreasing the temperature down to  $15^\circ\text{C}$ , kinetic traces collected for AncINS, AncFT and RLuc8 exhibited a triple-exponential decay of the fluorescence signal.

The two initial kinetic phases related to the binding of the substrate were fit with a double exponential equation (Eq. (9)) for each fluorescence trace (Supplementary Fig. 8). The dependence of the observed rates on the substrate concentration for both the fast and the slow kinetic phases were fit according to the induced-fit (IF) mechanism (Eqs. (10)–(12)) and the conformational selection (CS) mechanism (Eqs. (13)–(15)), assuming that  $k_{-1} \gg k_{-2}, k_{+2}$ <sup>57</sup>. The obtained values were used as initial estimates of the elementary rate constants describing the

binding process.

$$F = F_{SS} + A_1 \cdot e^{-k_{fast} \cdot t} + A_2 \cdot e^{-k_{slow} \cdot t} \quad (9)$$



$$k_{fast,IF} = k_{+1} \cdot [S] + k_{-1} \quad (11)$$

$$k_{slow,IF} = \frac{k_{+2} \cdot \frac{k_{+1}}{k_{-1}} \cdot [S]}{\frac{k_{+1}}{k_{-1}} \cdot [S] + 1} + k_{-2} \quad (12)$$



$$k_{fast,CS} = k_{+2} \cdot [S] + k_{-1} \quad (14)$$

$$k_{slow,CS} = \frac{k_{+1} \cdot \left( \frac{k_{+1}}{k_{-1}} \cdot k_{+2} \cdot [S] + k_{-2} \right)}{k_{+1} \cdot \frac{k_{+1}}{k_{-1}} \cdot k_{+2} \cdot [S] + k_{-2}} \quad (15)$$

In the case of Anc<sup>HLD</sup>-RLuc, the kinetic traces exhibited slow equilibration with low signal amplitudes, resulting in a limited precision of analysis (Supplementary Fig. 8). The data could be fit with only a single exponential equation (Eq. 16), and the dependence of the observed rate constant and the amplitude on the substrate concentration were fit according to a one-step binding mechanism (Eqs. (17)–(19)). Such a result pointed out that the rigidified ancestral enzyme shows no profound conformational flexibility and that only a simple substrate binding could be detected.

$$F = F_{SS} + A \cdot e^{-k_{obs} \cdot t} \quad (16)$$



$$k_{obs} = k_{+1} \cdot [S] + k_{-1} \quad (18)$$

$$A = \frac{A_{lim} \cdot [S]}{K_d + [S]} \quad (19)$$

The original raw kinetic data with residuals normalized by sigma values were subsequently analysed globally by numerical simulation using the KinTek Explorer software (KinTek Corporation, USA)<sup>52,55,56</sup> and following the same protocol as described in the steady-state kinetics section. The time-course of the fluorescence quenching upon substrate binding was described with Eqs. (20)–(22) for the induced-fit, conformational-selection, and one-step binding mechanisms, respectively. The estimates of the rate constants obtained by analytical fitting were used as starting points, and the final values were obtained by fitting the mechanism directly to the collected kinetic data. The quality of the best fits for different mechanisms was compared in terms of  $\chi^2$  value, “chi-by-eye” assessment, and confidence contour analysis assessing constraint and independency of the determined parameters. Based on this analysis, the correct mechanism of the substrate binding and the respective elementary rate constants were identified. Lower and upper limits for each parameter were derived for the  $\chi^2$  threshold of 0.95<sup>56</sup>.

$$F_{IF} = f \cdot \left( [E] + a \cdot [ES] + b \cdot [E^*S] \right) \quad (20)$$

$$F_{CS} = f \cdot \left( [E] + a \cdot [E^*] + b \cdot [E^*S] \right) \quad (21)$$

$$F_{one-step} = f \cdot \left( [E] + a \cdot [ES] \right) \quad (22)$$

**Hydrogen/deuterium exchange (HDX) mass spectrometry.** Enzyme samples were diluted with 100 mM phosphate buffer in H<sub>2</sub>O (pH 7.5) to prepare undeuterated controls and for peptide mapping, or with 100 mM phosphate buffer in D<sub>2</sub>O (pD 7.1) to prepare deuterated samples. The final concentrations of the enzyme samples used for analysis was 2  $\mu$ M. HDX was carried out at room temperature and was quenched after 10, 60 or 300 s by adding 1 M HCl in 1 M glycine with pepsin. Each sample was directly injected into an LC-system (UltiMate 3000 RSLCnano, Thermo Fisher Scientific, USA) with an immobilized nepenthesin enzymatic column (Affipro, CZ; 15  $\mu$ l bed volume, flow rate 20  $\mu$ l/min, 2% acetonitrile/0.05% trifluoroacetic acid). Peptides were trapped and desalted on-line on a peptide microtrap (Michrom Biosources, CA) for 3 min at a flow rate of 20  $\mu$ l/min. Next, the peptides were eluted onto an analytical column (Jupiter C18, 1.0  $\times$  50 mm, 5  $\mu$ m, 300  $\text{\AA}$ , Phenomenex, CA) and separated by linear gradient elution starting with 10% buffer A in buffer B and rising to 40% buffer A over 2 min. This was

followed by 31 min isocratic elution at 40% B. Buffers A and B consisted of 0.1% formic acid in water and 80% acetonitrile/0.08% formic acid, respectively. The immobilized nepenthesin column, trap cartridge, and analytical column were kept at 1  $^{\circ}$ C using a refrigerated system (Science Instruments and Software, CZ).

Mass spectrometric analysis was carried out using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, USA) with ESI ionization connected on-line to a robotic system based on the HTS-XT platform (CTC Analytics, Switzerland). The instrument was operated in a data-dependent mode for peptide mapping (HPLC-MS/MS). Each MS scan was followed by MS/MS scans of the three most intensive ions from both CID and HCD fragmentation spectra. Tandem mass spectra were searched using SequestHT against the cRAP protein database (ftp://ftp.thegpm.org/fasta/cRAP) containing the sequences of Rluc8, Anc<sup>HLD</sup>-RLuc and the studied mutants with the following search settings: mass tolerance for precursor ions of 10 ppm, mass tolerance for fragment ions of 0.6 Da, no enzyme specificity, two maximum missed cleavage sites and no-fixed or variable modifications. The false discovery rate at the peptide identification level was set to 1%. Sequence coverage was analysed with Proteome Discoverer version 1.4 (Thermo Fisher Scientific, USA) and graphically visualized with the MS Tools application<sup>58</sup>. Analysis of deuterated samples was done in HPLC-LC-MS mode with ion detection in the orbital ion trap. The MS raw files together with the list of peptides (peptide pool) identified with high confidence characterized by requested parameters (amino acid sequence of each peptide, its retention time, XCorr, and ion charge) were processed using HDExaminer version 2.2 (Sierra Analytics, CA). The software used to analyse protein and peptide behaviour, creates the uptake plots that measured peptide deuteration over time with the calculated confidence level.

**HDX mass spectrometry data analysis.** The acquired data were analysed as follows. The raw MS files together with the list of peptides (peptide pool) identified with high confidence characterized by requested parameters (retention time, XCorr, and charge) were processed using HDExaminer version 2.2 (Sierra Analytics, Modesto, CA). The software analysed the proteins' and peptides' behaviour and generated uptake data that were mapped to the proteins' amino acid sequences via the following procedure. Each residue was assigned the uptake data from any peptide solved with high confidence. Medium confidence peptides were also accepted for positions without previously assigned data. Low confidence peptides were rejected. The final uptake value (expressed as % of deuteration) assigned to each amino acid corresponded to the average of all assigned values for its position.

**MD simulations.** MD simulations were carried out on the crystal structures of RLuc8 (PDB ID 2PSF<sup>17</sup>), Anc<sup>HLD</sup>-RLuc (PDB ID 6G75<sup>14</sup>), AncINS (PDB ID 6S6E) and AncFT (PDB ID 6S97). Hydrogen atoms were added to the structures using the H++ web server<sup>59</sup>, at pH 7.5. Water molecules from the crystal structures, which did not overlap with the protonated structures, were retained. The systems were solvated using the solvate module of high-throughput molecular dynamics (HTMD) in a cubical water box of TIP3P water molecules so that all atoms were at least 10  $\text{\AA}$  from the surface of the box<sup>60</sup>. Cl<sup>-</sup> and Na<sup>+</sup> ions were added to neutralize the protein's charge and get a final concentration of 0.1 M. In all the simulations periodic boundary conditions were applied, the particle mesh Ewald method was used to treat interactions beyond a 9  $\text{\AA}$  cut-off, electrostatic interactions were suppressed >4 bond terms away from each other, and the smoothing and switching of van der Waals and electrostatic interactions were cut-off at 7.5  $\text{\AA}$ <sup>61</sup>. HTMD was used for adaptive sampling of the RMSD of the Ca atoms of residues 10–290. The production runs (20 ns) were started with the files resulting from the equilibration using settings from the final equilibration step. The trajectories were saved every 100 ps. The adaptive epochs were updated every 6 h to run with a minimum of five simulations and a maximum of 10 simulations for the total maximum of 110 epochs using TICA in 1 dimension<sup>62</sup>.

**Equilibration of systems for MD simulations.** The systems were equilibrated using the Equilibration\_v2 module of HTMD 1.23.4<sup>59</sup>. The system was first minimized using the conjugate-gradient method for 500 steps, after which the system was heated and minimized as follows (i) 500 steps (2 ps) of NVT heating, with the Berendsen barostat, to 310 K, with constraints on all heavy atoms of the protein; (ii) 2.5 ns of NPT equilibration, with the Langevin thermostat, with 1 kcal mol<sup>-1</sup>  $\text{\AA}$ <sup>-2</sup> constraints on all the heavy atoms of the protein; and (iii) 2.5 ns of NPT equilibration, with the Langevin thermostat, without constraints. During the equilibration simulations, holonomic constraints were applied to all hydrogen-heavy atom bond terms and the mass of hydrogen atoms was scaled by a factor of 4, enabling the use of a 4 fs timestep<sup>60,61</sup>. This approach transfers some mass from the heavy atom to the hydrogen to slow the fastest vibrations in the hydrogen atom, which is acceptable because the thermodynamic properties of biological systems are insensitive to the distribution of atomic masses<sup>63,64</sup>.

**Adaptive sampling.** Adaptive sampling here refers to the MD method that performs simulations in small sequential batches of 10 simulations. Before the next batch of simulations is submitted, the first batch is analysed considering a certain parameter (RMSD in this study) and the less sampled conformational regions are used to start a new batch. This way the method ensures that the sampled regions

are maximized. This sampling analysis is done using Markov state models making a list of metastable states distributed through the conformational space.

**Calculation of B-factors.** B-factors were calculated from all snapshots obtained from the MD simulations for the backbone atoms of each enzyme. The Metric Fluctuation tool from the HTMD package was used to calculate the RMSF, from which B-factors were calculated using Eq. (23).

$$B = \left[ \frac{8 * \pi^2}{3} \right] * \text{RMSF}^2 \quad (23)$$

The B-factors obtained from the MD simulations were standardized for each enzyme to enable comparisons between enzymes. This was done using Eq. (24).

$$Z = \frac{X - \mu}{\delta} \quad (24)$$

Here,  $Z$  is the standardized value,  $X$  is the original B-factor value,  $\mu$  is the average of the B-factors, and  $\delta$  is the sample standard deviation. The simulations were made into a simulation list using HTMD, water was filtered out, and crashed simulations with lengths below 20 ns were omitted<sup>59</sup>. The total simulation time required for the calculations was in excess of 4  $\mu$ s for each enzyme. The dynamics of the enzymes were evaluated based on the RMSD of the C $\alpha$  atoms of residues 10–290 to avoid problems due to the misleadingly high RMSD values of the dynamically free residues at each end of the protein. The data were clustered into 200 clusters using the MiniBatchKmeans algorithm. The implied timescale plot (based on a Markov state model with various lag times) was constructed to select a lag time for Markov model construction, using the RMSD of the full protein as a metric. Because the timescales mostly stabilized after 15 ns lag time, this value was used in the models to construct the 2 Markov states (open and closed). The final data were calculated from 1000 bootstrapping runs using 50% of the data.

**Tunnel calculations.** The tunnels were calculated with Caver 3.02<sup>23</sup> using 1000 random frames belonging to each state using Caver Analyst 2.0.31<sup>65</sup>, with Probe radius: 0.9 Å, clustering threshold: 5, shell depth: 5 Å and shell radius: 5 Å. For the Anc<sup>HLD</sup>-RLuc the starting point of the tunnel was defined by: ND2 of the Asn51 and OE1 of Glu142. For other enzymes, the starting point was defined by ND2 and OE1 in analogous residues.

**Cavities calculations.** The cavity calculations for the four proteins were carried out in the duplicates using the software Caver Analyst 2.0.31<sup>65</sup>. We have used 100 random snapshots to obtain the cavity volumes from the open conformations and 100 random snapshots from the closed conformations. A probe of 1.4 Å was used for the cavities and a maximum probe size of 3.0 Å was used to delimit the interface between solvent and cavity. The same parameters for the probes were used when calculating the cavity volume of the crystal structures. When two chains were present, the cavities were calculated for both chains. Cavities were also calculated for the minimized crystal structures, both chains when present.

**Crystallizations.** Diffraction-quality crystals of AncINS were obtained using the sitting-drop vapour diffusion technique in a 96-well crystallization plate (UVXPO 2 Lens Crystallization Plate, SWISSCI, Switzerland) at 20 °C after 5 days by mixing equal volumes of AncINS (4.5 or 10.5 mg/ml) with the reservoir solution from Morpheus crystallization screen condition C5, which consists of 0.09 M NPS additive mix (sodium nitrate, sodium phosphate dibasic, ammonium sulfate) and 0.1 M buffer system 2 (sodium HEPES, MOPS) at pH 7.5 with 30% v/v precipitant mix 1 (PEG 500 MME, PEG 20,000). The crystals were fished out and directly flash-frozen in liquid nitrogen for diffraction analysis at ESRF beamline ID23-1 in Grenoble (France).

AncFT crystals were obtained using the hanging-drop vapour diffusion technique in EasyXtal 15-well plates (Qiagen, Germany) with seeding at 20 °C after 2 days by mixing 0.5  $\mu$ l of seed, 2  $\mu$ l of AncFT (9.2 mg/ml), and 1  $\mu$ l of reservoir solution (0.2 M sodium acetate, 0.1 M Tris pH 8.5, 19% PEG 3350). Crystals were cryo-protected with 20% glycerol and flash-cooled in liquid nitrogen. Crystallographic data were collected at cryogenic temperature at SLS beamline PXIII in Villigen (Switzerland), with wavelength of 1 Å.

Prior to the co-crystallization screening, the purified protein RLuc8-W121F/E144Q was concentrated to 6.0 or 8.5 mg/ml and mixed with 3–5 molar excess of coelenterazine (stock solution of coelenterazine was 11.8 mM in isopropanol). The mixture was incubated 1 h at 4 °C, and then precipitated material was removed by centrifugation (13,000 rpm/10 min/4 °C), and the supernatant containing enzyme-ligand complexes was directly crystallized. The crystallization was performed in Easy-Xtal 15-well crystallization plates in a hanging drop vapour diffusion technique, where enzyme-ligand drops (typically 1.5  $\mu$ l) were mixed with the reservoir solution (2.2 M NaHPO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub>, 0.1 M sodium acetate pH = 4.5) in the ratio 1:1 and equilibrated against 500  $\mu$ l of the reservoir solution. Diffraction quality crystals of RLuc8-W121F/E144Q were obtained at 20 °C after 3–5 days. All crystals used for X-ray diffraction analysis were flash-frozen in liquid nitrogen in the corresponding reservoir solutions supplemented with 20–25% glycerol. All data obtained in this project were collected at 100 K on SLS beamline PX3 (Villigen, Switzerland).

**Structure determination, model building and refinement.** The crystallographic data were processed and scaled using XDS<sup>66</sup> and Aimless<sup>67</sup>. Structures of AncINS and AncFT were solved by molecular replacement using Anc<sup>HLD</sup>-RLuc (PDB ID 6G75<sup>14</sup>) as a search model with the help of Phaser<sup>68</sup> as implemented in the Phenix package<sup>69</sup>. The structure of RLuc8-W121F/E144Q was solved by molecular replacement with Phenix2 using the RLuc8 structure (PDB ID 2PSD) as a search model. Multiple cycles of automated refinement were performed in the phenix.refine programme<sup>69</sup> and manual model building was performed in Coot<sup>70</sup>. The final models were validated using tools provided in Coot<sup>70</sup> and Molprobity<sup>71</sup>. Ramachandran favoured 95.3%, 95.9% and 96.6%, Ramachandran allowed 4.4%, 3.8% and 3.4%, Ramachandran outliers 0.3%, 0.3% and 0% for AncINS, AncFT and RLuc8-W121F/E144Q, respectively. Structural data were graphically visualized with PyMol Molecular Graphics System, version 1.8 (Schrödinger, LLC). Atomic coordinates and structure factors for AncINS, AncFT and RLuc8-W121F/E144Q were deposited in the Protein Data Bank under the codes 6S6E, 6S97, and 6YN2, respectively.

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

## Data availability

Atomic coordinates and experimental data have been deposited in the Protein Data Bank ([www.wwpdb.org](http://www.wwpdb.org)). They are publicly available for accession codes 6S6E and 6S97; 6YN2 will be released upon article publication. The validation reports are provided in Supplementary Notes 7, 8 and 9. The primary data from microfluidics are available in figshare with the identifier <https://doi.org/10.6084/m9.figshare.14453700.v1>. Source data are provided with this paper.

Received: 3 August 2020; Accepted: 27 April 2021;  
Published online: 14 June 2021

## References

- Baier, F., Copp, J. N. & Tokuriki, N. Evolution of enzyme superfamilies: comprehensive exploration of sequence–function relationships. *Biochemistry* **55**, 6375–6388 (2016).
- Tyzack, J. D., Furnham, N., Sillitoe, I., Orengo, C. M. & Thornton, J. M. Understanding enzyme function evolution from a computational perspective. *Curr. Opin. Struct. Biol.* **47**, 131–139 (2017).
- Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.* **57**, 4143–4148 (2018).
- Tóth-Petróczy, Á. & Tawfik, D. S. Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol. Biol. Evol.* **30**, 761–771 (2013).
- Hochberg, G. K. A. & Thornton, J. W. Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).
- Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
- Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).
- Zou, T., Risso, V. A., Gavira, J. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* **32**, 132–143 (2014).
- Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrected precambrian  $\beta$ -lactamases. *J. Am. Chem. Soc.* **135**, 2899–2902 (2013).
- Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. Evolution of the protein repertoire. *Science* **300**, 1701 (2003).
- Pascarella, S. & Argos, P. Analysis of insertions/deletions in protein structures. *J. Mol. Biol.* **224**, 461–471 (1992).
- Emond, S. et al. Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. *Nat. Commun.* **11**, 1–14 (2020).
- Skamaki, K. et al. In vitro evolution of antibody affinity via insertional scanning mutagenesis of an entire antibody variable region. *Proc. Natl Acad. Sci. USA* **117**, 27307–27318 (2020).
- Chaloupkova, R. et al. Light-emitting dehalogenases: reconstruction of multifunctional biocatalysts. *ACS Catal.* **9**, 4810–4823 (2019).
- Koudelakova, T. et al. Haloalkane dehalogenases: biotechnological applications. *Biotechnol. J.* **8**, 32–45 (2013).
- Lorenz, W. W., McCann, R. O., Longiaru, M. & Cormier, M. J. Isolation and expression of a cDNA encoding *Renilla reniformis* luciferase. *Proc. Natl Acad. Sci. USA* **88**, 4438–4442 (1991).

17. Loening, A. M., Fenn, T. D. & Gambhir, S. S. Crystal structures of the luciferase and green fluorescent protein from *Renilla reniformis*. *J. Mol. Biol.* **374**, 1017–1028 (2007).
18. Woo, J., Howell, M. H. & Arnim, A. G. Structure–function studies on the active site of the coelenterazine-dependent luciferase from *Renilla*. *Protein Sci.* **17**, 725–735 (2008).
19. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869 (2006).
20. Loening, A. M., Fenn, T. D., Wu, A. M. & Gambhir, S. S. Consensus guided mutagenesis of *Renilla* luciferase yields enhanced stability and light output. *Protein Eng. Des. Sel.* **19**, 391–400 (2006).
21. Bradshaw, R. T. et al. Neurotransmitter transporter conformational dynamics using HDX-MS and molecular dynamics simulation. *Biophys. J.* **114**, 207a (2018).
22. Yeh, H.-W. & Ai, H.-W. Development and applications of bioluminescent and chemiluminescent reporters and biosensors. *Annu. Rev. Anal. Chem. Palo Alto Calif.* **12**, 129–150 (2019).
23. Chovancova, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8**, e1002708 (2012).
24. Kipnis, Y., Dellus-Gur, E. & Tawfik, D. S. TRINS: a method for gene modification by randomized tandem repeat insertions. *Protein Eng. Des. Sel.* **25**, 437–444 (2012).
25. Jones, D. D. Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1  $\beta$ -lactamase to an amino acid deletion. *Nucleic Acids Res.* **33**, e80–e80 (2005).
26. Fujii, R., Kitaoka, M. & Hayashi, K. in *Directed Evolution Library Creation: Methods and Protocols* (eds Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 151–158 (Springer, 2014).
27. Jones, D. D., Arpino, J. A. J., Baldwin, A. J. & Edmundson, M. C. in *Directed Evolution Library Creation: Methods and Protocols* (eds Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 159–172 (Springer, 2014).
28. Obexer, R. et al. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
29. Khare, S. D. et al. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol.* **8**, 294 (2012).
30. Rockah-Shmuel, L. & Tawfik, D. S. Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res.* **40**, 11627–11637 (2012).
31. Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z. & Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).
32. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine learning in enzyme engineering. *ACS Catal.* **10**, 1210–1223 (2020).
33. Lourenço, J. M., Esteves da Silva, J. C. G. & Pinto da Silva, L. Combined experimental and theoretical study of Coelenterazine chemiluminescence in aqueous solution. *J. Lumin.* **194**, 139–145 (2018).
34. Magalhães, C. M., Esteves da Silva, J. C. G. & Pinto da Silva, L. Comparative study of the chemiluminescence of coelenterazine, coelenterazine-e and Cypridina luciferin with an experimental and theoretical approach. *J. Photochem. Photobiol. B* **190**, 21–31 (2019).
35. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
36. Krefß, N., Halder, J. M., Rapp, L. R. & Hauer, B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Energy Mech. Biol.* **47**, 109–116 (2018).
37. Nestl, B. M. & Hauer, B. Engineering of flexible loops in enzymes. *ACS Catal.* **4**, 3201–3211 (2014).
38. Park, H.-S. et al. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535–538 (2006).
39. Afriat-Jurnou, L., Jackson, C. J. & Tawfik, D. S. Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling. *Biochemistry* **51**, 6047–6055 (2012).
40. Bornscheuer, U. T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
41. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
42. Emond, S. & Hollfelder, F. TRIAD: a transposition-based approach for gene mutagenesis by random short in-frame insertions and deletions for directed protein evolution. *Protoc. Exch.* <https://doi.org/10.21203/rs.3.pex-1448/v1> (2021).
43. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
44. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
45. Holloway, P., Trevors, J. T. & Lee, H. A colorimetric assay for detecting haloalkane dehalogenase activity. *J. Microbiol. Methods* **32**, 31–36 (1998).
46. Buryška, T. et al. Controlled oil/water partitioning of hydrophobic substrates extending the bioanalytical applications of droplet-based microfluidics. *Anal. Chem.* **91**, 10008–10015 (2019).
47. Joosten, R. P. et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–D419 (2011).
48. Bakan, A. et al. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinforma. Oxf. Engl.* **30**, 2681–2683 (2014).
49. Wold, S. Validation of QSAR's. *Quant. Struct. -Act. Relatsh.* **10**, 191–193 (1991).
50. Wold, S., Johansson, E. & Cocchi, M. in *3D QSAR in Drug Design. Theory, Methods, and Applications* (ed. Kubinyi, H.) 523–550 (ESCOM Science Publisher, 1993).
51. Wold, S. & Dunn, W. J. Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comput. Sci.* **23**, 6–13 (1983).
52. Johnson, K. A. in *Methods in Enzymology* Vol. 467 (eds Johnson, M. L. & Brand, L.) Ch. 23, 601–626 (Academic Press, 2009).
53. Johnson, K. A. New standards for collecting and fitting steady state kinetic data. *Beilstein J. Org. Chem.* **15**, 16–29 (2019).
54. O'Kane, D. J. & Lee, J. Absolute calibration of luminometers with low-level light standards. *Methods Enzymol.* **305**, 87–96 (2000).
55. Johnson, K. A., Simpson, Z. B. & Blom, T. Global Kinetic Explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 (2009).
56. Johnson, K. A., Simpson, Z. B. & Blom, T. FitSpace Explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem.* **387**, 30–41 (2009).
57. Bagshaw, C. R. *Biomolecular Kinetics: A Step-by-Step Guide* (CRC Press, 2017).
58. Kavan, D. & Man, P. MSTools—Web based application for visualization and presentation of HXMS data. *Hydrog. Exch. Mass Spectrom.* **302**, 53–58 (2011).
59. Doerr, S., Harvey, M. J., Noé, F. & De Fabritiis, G. HTMD: high-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.* **12**, 1845–1852 (2016).
60. Harvey, M. J., Giuipponi, G. & Fabritiis, G. D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).
61. Harvey, M. J. & De Fabritiis, G. An implementation of the smooth particle mesh ewald method on GPU hardware. *J. Chem. Theory Comput.* **5**, 2371–2377 (2009).
62. Naritome, Y. & Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.* **134**, 065101 (2011).
63. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).
64. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
65. Jurcik, A. et al. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**, 3586–3588 (2018).
66. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
67. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution?. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–1214 (2013).
68. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
69. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods Struct. Proteomics* **55**, 94–106 (2011).
70. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
71. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).

## Acknowledgements

The authors would like to express thanks to Dr. Chris Badenhorst (University Greifswald) for his very helpful comments on the manuscript. This work was supported by grants from the Czech Ministry of Education (LQ1605, CZ.02.1.01/0.0/0.0/16\_013/0001761, LM2015051, LM2015047, LM2015055), the Grant Agency of the Czech Republic (17-24321S), the Ministry of Health of the Czech Republic MH CZ-DRO MMCI, 00209805 and the European Union (720776, 722610, 814418). J.P.-I. is supported by Postdoc/MUNI (CZ.02.2.69/0.0/16\_027/0008360), M.T. and M.V. by Brno Ph.D. Talent. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 792772. M.M. acknowledges financial support from GAMU of the Masaryk University (MUNI/H/1561/2018). Computational resources were provided by CERIT-SC and Meta Centrum (LM2015042, LM2015085). F.H. is an H2020 ERC Advanced Investigator [695669] and work in Cambridge was supported by grant BB/L002469/1 from the BBSRC (Biological Sciences Research Council). We thank members of the

European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) synchrotrons for the use of their beamline facilities and for help during data collection. Core Facility Biomolecular Interactions and Crystallization of CEITEC MU is gratefully acknowledged for obtaining scientific data presented in this paper. A.S. is on leave from the Institute of Chemistry, Slovak Academy of Sciences, 845 38 Bratislava, Slovak Republic. We are grateful to Klaudia Chmelova (Masaryk University) for her kind assistance during X-ray data collection and Dr. Stanislav Mazurenko (Masaryk University, Brno) for help with mathematical analysis.

### Author contributions

V.D.L. and S.E. constructed libraries and performed manual screening; A.S., M.T., R.C., V.D.L., M.V. and M.M. conducted biochemical and biophysical characterization of protein variants; A.S. and M.D. performed robotic screening; A.S. and M.M. crystallized proteins, collected X-ray data and determined structures; A.S. and M.M. designed and performed experiments with mammalian cell lysates; G.P.P., J.P.-I. and D.B. performed and analysed MD simulations; D.P., M.T. and Z.P. performed steady-state and transient kinetic experiments and analysed data; L.H. conducted HDX-MS experiments and analysed data; J.D. and J.P.-I. conducted bioinformatics analysis; J.D., F.H. and U.B.T. designed the study and contributed to data interpretation. All authors contributed to the writing of the manuscript and preparation of tables and figures.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23450-z>.

**Correspondence** and requests for materials should be addressed to F.H., U.T.B. or J.D.

**Peer review information** *Nature Communications* thanks Kirill Alexandrov, Vitor Pinheiro and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021