De novo active sites for resurrected Precambrian enzymes

Valeria A. Risso 1, Sergio Martinez-Rodriguez 1, Adela M. Candel 1, Dennis M. Krüger 2, David Pantoja-Uceda 3, Mariano Ortega-Muñoz 4, Francisco Santoyo-Gonzalez 4, Eric A. Gaucher 5, Shina C.L. Kamerlin 2, Marta Bruix 3, Jose A. Gavira 6 & Jose M. Sanchez-Ruiz 1

Protein engineering studies often suggest the emergence of completely new enzyme functionalities to be highly improbable. However, enzymes likely catalysed many different reactions already in the last universal common ancestor. Mechanisms for the emergence of completely new active sites must therefore either plausibly exist or at least have existed at the primordial protein stage. Here, we use resurrected Precambrian proteins as scaffolds for protein engineering and demonstrate that a new active site can be generated through a single hydrophobic-to-ionizable amino acid replacement that generates a partially buried group with perturbed physico-chemical properties. We provide experimental and computational evidence that conformational flexibility can assist the emergence and subsequent evolution of new active sites by improving substrate and transition-state binding, through the sampling of many potentially productive conformations. Our results suggest a mechanism for the emergence of primordial enzymes and highlight the potential of ancestral reconstruction as a tool for protein engineering.

1 Departamento de Quimica Fisica, Facultad de Ciencias University of Granada, 18071 Granada, Spain. 2 Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden. 3 Departamento de Quimica Fisica Biologica, Instituto de Quimica Fisica Rocasolano, CSIC, c/Serrano 119, 28006-Madrid, Spain. 4 Departamento de Quimica Organica, Facultad de Ciencias University of Granada, 18071 Granada, Spain. 5 School of Biology, School of Chemistry and Biochemistry, Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30322, USA. 6 Laboratorio de Estudios Cristalograficos, Instituto Andaluz de Ciencias de la Tierra, CSIC-University of Granada Avenida de la Palmeras 4, Granada, 18100 Armilla, Spain. Correspondence and requests for materials should be addressed to S.C.L.K. (email: kamerlin@icm.uu.se) or to J.M.S.-R. (email: sanchezr@ugr.es).
The generation of completely new active sites capable of enzyme catalysis is, arguably, one of the most fundamental unsolved problems in protein science. Rational design approaches to this problem have often used complex computational methods, have targeted simple model reactions and have typically led to low levels of catalysis. These studies would seem to suggest, therefore, the unlikelihood of the emergence of completely new active sites in non-catalytic scaffolds. Certainly, most modern enzyme functions likely evolved from previously existing functionalities. On the other hand, most biochemical processes are extremely slow in the absence of enzymes and specialized enzymes are likely to have catalysed many different reactions already in the last universal common ancestor. It could be inferred from this that efficient mechanisms for the emergence and subsequent evolution of completely new enzyme functionalities must exist or, at least, that they must have existed at the primordial protein stage. However, little is known about such mechanisms.

Buried and partially buried ionizable groups with perturbed properties often play essential catalytic roles in modern enzymes. Single hydrophobic-to-ionizable residue mutations that generate properties likely linked to enhanced conformational flexibility could have assisted the emergence of the new enzyme functionalities and its subsequent evolution by facilitating substrate and transition-state binding, through sampling a greater number of potentially productive conformations. Many years ago, Jensen proposed in a highly influential article that primordial enzymes were capable of catalysing a diversity of reactions. It is conceivable that the conformational flexibility that is likely linked to such a broad generalist nature may have facilitated the emergence of new enzyme functionalities in the first place.

Here, we explore and test these notions using resurrected Precambrian β-lactamases as scaffolds for the engineering of completely new active sites. To date, only a handful of systems have been studied using ancestral resurrection, and only a few of these resurrection efforts targeted ‘old’ (3 billion years) phylogenetic nodes (Fig. 2 of ref. 13). Of these systems, resurrected Precambrian β-lactamases have been thoroughly characterized in terms of their structure, function and stability. These putative ancestral proteins have been shown to be highly stable and to efficiently degrade several antibiotics. Previous and current computational analyses, as well as new NMR relaxation studies reported here, support that this substrate promiscuity is linked to enhanced conformational flexibility. The broad substrate scope of resurrected Precambrian β-lactamases may reflect the wide variety of substances these enzymes had to hydrolyse. We do not claim, therefore, that they are necessarily at Jensen’s ancestral generalist stage, although, strictly speaking, this possibility should not be ruled out. In any case, the available resurrected Precambrian β-lactamases should provide an adequate model with which to address the role of ancestral conformational flexibility in the emergence of new enzyme functionalities. More generally, we have recently reviewed several arguments and recent publications that support that promiscuity may be a common outcome of ancestral protein resurrection.

Here, we use carefully selected systems that span a vast region of the sequence space of both ancestral and modern β-lactamases, and probe the enzymatic features that allow for the emergence of a non-natural activity in these enzymes, as well as why it only appears in a specific snapshot of evolutionary time. We demonstrate the role of conformational flexibility in allowing for the emergence of new enzymatic functions, as well as its importance in the subsequent evolvability of the enzyme. Finally, our data highlights the potential of ancestral reconstruction as a tool for protein engineering by providing far more powerful evolutionary starting points than can be obtained from modern enzymes.

Results

Selected model systems. In the present study, we have used the proteins encoded by the most probabilistic sequences at six Precambrian phylogenetic nodes in the evolution of class A β-lactamases (Fig. 1). The reconstructed sequences and the procedure used to obtain them have been described in detail in Supporting Information of ref. 11. These proteins display large sequence differences between themselves (Supplementary Table 1) and they are properly folded, highly stable, active and share the β-lactamase fold. In addition, we have also used here proteins encoded by alternative sequences at the GNCA node (common ancestor of Gram-negative bacteria). As is customary in this field, these alternative sequences were derived from a random weighted sampling of the posterior probability distribution. They differ from the most probabilistic sequence at 8–20 positions (Supplementary Tables 2 and 3). Finally, for comparison, we have also used 10 modern β-lactamases (Fig. 1) that have been considered in the literature to be archetypical examples of β-lactamases and are well characterized in terms of their structure and function. These modern proteins provide a fair representation of the β-lactamases from the several bacterial taxa (Fig. 1) and they span modern β-lactamase sequence space to a substantial extent, as they show limited sequence identity between themselves (Supplementary Table 4). Note also their limited sequence identity with the putative ancestral β-lactamases studied here (Supplementary Table 5). Overall, we explore in this work a vast region of the sequence space of both ancestral and modern β-lactamases.

We have selected the Kemp elimination reaction (Fig. 2) as our primary target for the reaction to be catalysed by the generated new active site, for several reasons. First, it provides a simple activated model for proton abstraction from carbon, which is a fundamental chemical process that underlies many biochemical reactions. Completely new active sites are expected to display low catalysis levels (high catalysis levels would be the outcome of subsequent evolution) and their emergence is best probed by activated substrates. Second, Kemp elimination is a non-natural reaction (unknown in biological organisms). No natural enzyme has evolved, therefore, to catalyse this reaction and studies on engineered Kemp eliminases are unlikely to be compromised by contamination from natural enzymes. Following from this, Kemp elimination has been often used as a benchmark for rational enzyme design. Therefore, use of the same system allows for a quantitative comparison of the catalytic efficiencies of the designed constructs presented in this work to those generated in previous enzyme design studies.

Generation of a new active site for Kemp elimination. The possibility that slow conformational changes (on the micro-seconds to seconds timescale) play roles in enzyme catalytic cycles has been proposed and explored. Consequently, we used NMR relaxation determinations (see Methods section and Supplementary Methods) on the β-lactamase encoded by the most probabilistic sequence at the GNCA node (GNCAmp-β-lactamase) to guide our design of a new active site. This putative ancestral protein displays a large number of residues with conformational contribution to the relaxation rates determined by NMR (see the comparison with the modern TEM-1 β-lactamase shown in Fig. 3). A large accumulation of
such residues is visually apparent in the region encompassing helix h1 (residues 26–41), helix h11 (residues 271–290) and the loops 225–229 and 252–257. Residue 229 within this region appears as a suitable target for the generation of a new active site through a hydrophobic-to-ionizable residue replacement. A substantially buried and highly conserved tryptophan residue is present at position 229 in both modern and reconstructed Precambrian β-lactamases. The indole side chain of tryptophan has a shape similar to that of the Kemp substrate. Precisely because of this shape congruence, replacement of W229 to a new residue with basic properties will not generate a new active site capable of catalysing the Kemp elimination unless, of course, conformational rearrangements occur in the enzyme to be able to avoid steric clashes between the substrate and the new residue.

We found that a simple W229D replacement leads to substantial levels of Kemp elimination activity in the β-lactamases at all ancestral nodes studied here (Supplementary Table 6), with the only exception of the evolutionary recent ENCA node (Fig. 1). By substantial activity, we mean that the observed levels of Kemp elimination activity were clearly distinguishable (and much higher in most cases) than the background (enzyme-free) levels (Supplementary Fig. 1). In contrast, the W229D variants of all 10 modern β-lactamases studied led to levels of Kemp elimination activity that, even at protein concentrations of about 20 μM, could be barely distinguished from the background levels (Supplementary Fig. 1). It is important to note that the W229D variants of all 10 modern β-lactamases studied did show antibiotic degradation activity (linked to the natural active site) at nM concentrations (Supplementary Fig. 2). Therefore, their lack of Kemp elimination activity cannot be attributed to the ‘disruptive’ W229D mutation preventing their folding, which is anyhow a problem that is less likely to arise with the highly stable ancestral β-lactamases. It is worth noting here that enhanced stability is a common outcome of Precambrian protein resurrection11,18–20, likely linked to the thermophilic nature of early life.

A wide diversity of experimental results confirm (or are consistent with) the main features of the designed approach used. Specifically, X-ray crystallography in the presence of 5(6)-nitrobenzotriazole, a known transition-state analogue of the Kemp elimination reaction (Fig. 4) and inhibition by this analogue (Supplementary Fig. 3) support that Kemp elimination does occur at the site generated by the W229D replacement. The catalytic role of the aspartate at position 228 does not have a catalytic role, as the replacement of D228 with A does not substantially impair the activity that, even at protein concentrations of about 20 μM, could be barely distinguished from the background levels (Supplementary Fig. 2). Therefore, their lack of Kemp elimination activity cannot be attributed to the ‘disruptive’ W229D mutation preventing their folding, which is anyhow a problem that is less likely to arise with the highly stable ancestral β-lactamases. It is worth noting here that enhanced stability is a common outcome of Precambrian protein resurrection11,18–20, likely linked to the thermophilic nature of early life.

Specifically, X-ray crystallography in the presence of 5(6)-nitrobenzotriazole, a known transition-state analogue of the Kemp elimination reaction (Fig. 4) and inhibition by this analogue (Supplementary Fig. 3) support that Kemp elimination does occur at the site generated by the W229D replacement. The catalytic role of the aspartate at position 229 is further confirmed by mutational studies: the W229G variant shows negligible activity and the nearby aspartate at position 228 does not have a catalytic role, as the replacement of D228 with A does not substantially impair the activity (Supplementary Table 6). The catalytic role of the aspartate at position 229 is also confirmed by the pH dependence of the catalysis, which is consistent with the raised pK value expected for an aspartate residue in a hydrophobic environment (Fig. 5).

Following from this, 3D-structure determination in the presence of 5(6)-nitrobenzotriazole, a known transition-state analogue of the Kemp elimination reaction (Fig. 4) and inhibition by this analogue (Supplementary Fig. 3) support that Kemp elimination does occur at the site generated by the W229D replacement. The catalytic role of the aspartate at position 229 is further confirmed by mutational studies: the W229G variant shows negligible activity and the nearby aspartate at position 228 does not have a catalytic role, as the replacement of D228 with A does not substantially impair the activity (Supplementary Table 6). The catalytic role of the aspartate at position 229 is also confirmed by the pH dependence of the catalysis, which is consistent with the raised pK value expected for an aspartate residue in a hydrophobic environment (Fig. 5).

![Figure 1](https://example.com/f1.png) **Figure 1 | Ancestral and modern β-lactamases used as scaffolds in this work.** Schematic representation of the phylogenetic tree for class A β-lactamases11. The ancestral nodes studied in this work correspond to the common ancestors (CA) of Firmicutes (FCA), Actinobacteria and Firmicutes (AFCA), Enterobacteria (ENCA), Gammaproteobacteria (GPBCA), various Gram-negative bacteria (GNCA) and various Gram-positive and Gram-negative bacteria (PNCA). The proteins encoded by the common ancestors (CA) of Firmicutes (FCA), Actinobacteria and Firmicutes (AFCA), Enterobacteria (ENCA), Gammaproteobacteria (GPBCA), various Gram-negative bacteria (GNCA) and various Gram-positive and Gram-negative bacteria (PNCA). The proteins encoded by the most probabilistic sequences at these nodes11 were prepared and used as scaffolds for engineering. In addition, seven alternative reconstructions at the GNCA node (Supplementary Tables 2 and 3) were also prepared; these are labelled GNCA1 to GNCA7, with the most probabilistic sequence at the node being labelled GNCA7. β-lactamases from 10 modern organisms are also studied in this work and these organisms are shown at the right.

![Figure 2](https://example.com/f2.png) **Figure 2 | Reactions and compounds studied in this work.** (a) The Kemp elimination of 5-nitrobenzoxazole (5-nitro-benzo[d]isoxazole). A schematic transition state structure is shown here; (b) tryptophan; (c) 5(6)-nitrobenzotriazole (a transition-sate analogue) and (d) indole (the tryptophan side chain). Finally, (e) the hydrolysis of p-nitrophenyl acetate is also shown here.
the transition-state analogue (Fig. 4) confirms the role of conformational flexibility in the generation of a new function, as transition-state binding is shown to rely on conformational rearrangements. Specifically, the bound transition-state analogue is displaced with respect to the position originally occupied by the tryptophan 229, as required by the presence of an aspartate residue at position 229 in the active variants. Such a displacement is made possible by a shift of the h11 α-helix (and the concomitant shift of the substantially solvent-exposed h1 α-helix). We note also that the catalytic efficiency for the Kemp elimination in the engineered ancestral proteins correlates with the transition-state analogue binding constants derived from inhibition experiments (Supplementary Fig. 4), supporting that catalysis is linked to transition-state stabilization. Finally, catalysis of Kemp elimination is enhanced by an additional F290W amino replacement in the neighbourhood of position 229, which likely stabilizes the transition state through a face-to-edge interaction with the new tryptophan residue (Fig. 4).

Comparison with previous rational designs. Kemp elimination is a non-natural reaction that is unknown to biological organisms. No enzyme is, therefore, expected to have evolved to catalyse this reaction. Kemp eliminases are engineered through the introduction of transition-state analogues binding to the enzyme active site. The transition-state analogue (Fig. 4) confirms the role of conformational flexibility in the generation of a new function, as transition-state binding is shown to rely on conformational rearrangements. Specifically, the bound transition-state analogue is displaced with respect to the position originally occupied by the tryptophan 229, as required by the presence of an aspartate residue at position 229 in the active variants. Such a displacement is made possible by a shift of the h11 α-helix (and the concomitant shift of the substantially solvent-exposed h1 α-helix). We note also that the catalytic efficiency for the Kemp elimination in the engineered ancestral proteins correlates with the transition-state analogue binding constants derived from inhibition experiments (Supplementary Fig. 4), supporting that catalysis is linked to transition-state stabilization. Finally, catalysis of Kemp elimination is enhanced by an additional F290W amino replacement in the neighbourhood of position 229, which likely stabilizes the transition state through a face-to-edge interaction with the new tryptophan residue (Fig. 4).
reaction and, consequently, Kemp elimination activity is not expected among natural proteins. Indeed, Hilvert and coworkers\(^\text{21}\) found no detectable Kemp eliminase activity in calmodulin, barnase, lysozyme, trypsin, chymotrypsinogen and chymotrypsin. Furthermore, Tawfik and coworkers\(^\text{16}\) screened the ASKA library (~4,300 clones) for natural and non-natural activities and found only two enzymes that could catalyze Kemp elimination, although these appeared to be promiscuous activities. On the other hand, serum albumins can catalyse Kemp elimination\(^\text{22}\) with a turnover number that is about 10\(^{-2}\) s\(^{-1}\) at neutral pH, but that approaches 10 s\(^{-1}\) at pH ~10. However, serum albumins, even in the absence of bound metals, can catalyse a remarkable diversity of non-natural reactions\(^\text{23}\), due to hydrophobic pockets, which can bind different substrates and to the presence of lysine residues with catalytic properties within those pockets.

The capability of albumins to catalyse the Kemp elimination reaction (among several other non-natural reactions) suggests that the scarcity of natural enzyme catalysts for such a simple reaction is simply due to the lack of a selective pressure to generate Kemp elimination activity during evolution. On the other hand, rational design efforts to engineer artificial Kemp eliminases have resulted in limited success\(^\text{1}\), despite the simplicity of the targeted reaction. Indeed, comparison with these previous Kemp eliminase designs (Fig. 6) provides further evidence that the catalytic efficiencies \((k_{\text{cat}}/K_m)\) values we have generated in the ancestral backgrounds used is actually similar or clearly above the results previously reported using minimalist design\(^\text{24,25}\). The upper end of the range is above the results previously reported using more complex design approaches\(^\text{26,27}\) (iterative design and design based on Rosetta), which also involved large numbers of mutations to reach those efficiencies, and it is less than two orders of magnitude below the catalytic efficiency for the best Kemp eliminase reported for date\(^\text{28}\) (which is the outcome of 17 rounds of directed evolution from an iterative design background).

Regarding the turnover numbers, the \(k_{\text{cat}}\) values obtained for the W229D/F290W variants of the ancestral \(\beta\)-lactamase scaffolds are up to approximately seven orders of magnitude above the rate of the uncatalysed reaction (Fig. 6). For comparison, the best artificial Kemp eliminase reported to date\(^\text{28}\) displayed a approximately nine orders of magnitude enhancement, but, as noted above, this was obtained as a result of 17 rounds of directed evolution on a designed background with already substantial activity. Furthermore, the upper end of the range spanned by our \(k_{\text{cat}}\) values is clearly above the reported \(k_{\text{cat}}\) values to be calculated. However, lower limit estimates of \(k_{\text{cat}}\) can indeed be derived from suitable analysis of linear Michaelis plots (see Supplementary Fig. 5 for details). These estimated lower limit values for our best Kemp eliminase, GNCA\(_4\)-W229D/F290W, are shown as dashed lines. \(K_m\) values for this variant (shown in c) appear to be essentially pH-independent.

---

**Figure 5 | pH dependence of the de novo Kemp eliminases.** (a) Plot of catalytic efficiency \((k_{\text{cat}}/K_m)\) versus pH. (b) Plot of turnover number \((k_{\text{cat}})\) versus pH. (c) Plot of the Michaelis constant \((K_m)\) versus pH. The data shown are derived from the fitting of the Michaelis-Menten equation to the experimental profiles of rate versus substrate concentration. Error bars stand for the s.d.’s of the fitting parameters as provided by the fitting program used (Igor Pro 6.37). The colour of the data points refers to the \(\beta\)-lactamase variant studied, specifically: GNCA\(_{MP}\)-W229D (green); GNCA\(_3\)-W229D (blue); GNCA\(_4\)-W229D/F290W (red). The continuous lines in (a,b) represent the best fits of equation (1) to the experimental data. The pK\(_a\) values determined from such fits are shown alongside the lines, together with the standard errors derived from the fittings. The high pH limiting values for our best Kemp eliminase, GNCA\(_4\)-W229D/F290W, are shown as dashed lines. \(K_m\) values for this variant (shown in c) appear to be essentially pH-independent.
Finally, it must be noted that catalysis of Kemp elimination by carboxylic acids is strongly accelerated in aprotic solvents. The acetate ion in acetonitrile is in fact an excellent catalyst of the Kemp elimination with a reported second-order rate constant of 2,800 M$^{-1}$ s$^{-1}$, a value that has been used as a metric to judge the catalytic efficiency of artificial Kemp eliminases. Unlike previous rationally designed Kemp eliminases, our best eliminase displays a maximum catalytic efficiency ($\sim$5,500 M$^{-1}$ s$^{-1}$, Fig. 5a) that exceeds the acetate in acetonitrile level.

**Additional activities of the engineered ancestral enzymes.** We used our best eliminase (the W229D/F290W variant of the alternative GNCA4 reconstruction at the GNCA node), to test whether the introduction of the new active site affects the catalysis at the ‘old’ natural active site (catalytic residue S70, located at about 23 Å from position 229). β-Lactamas encoded by reconstructed sequences corresponding to the GNCA node have been previously shown to be able to degrade a variety of antibiotics, including penicillin and third-generation antibiotics with efficiencies similar to that of a modern average enzyme (by contrast, the modern TEM-1 β-lactamase is a specialist enzyme that displays high catalytic efficiency with penicillin and a substantially lower efficiency with third-generation antibiotics). We therefore determined the Michaelis–Menten parameters for the degradation of a penicillin antibiotic (benzylpenicillin) and one third-generation antibiotic (cefotaxime) catalysed by GNCA4-W229D/F290W (Fig. 7). We found levels of catalysis for antibiotic degradation similar to those previously reported for the GNCA$_{\text{MP}}$ β-lactamase (Supplementary Table 7).

We also used our best eliminase to explore the potential promiscuity at the generated new active site. Our results show that the GNCA4-W229D/F290W β-lactamase does in fact also catalyse the hydrolysis of p-nitrophenyl acetate (Figs 2 and 7 and Supplementary Fig. 6), a substrate commonly used to assess esterase activity. However, lactam hydrolysis and ester hydrolysis are chemically similar and the natural (antibiotic degradation) active site could therefore in principle contribute to the observed esterase activity. Nevertheless, our results (Fig. 7) indicate a minor natural-site contribution that only becomes apparent when the de novo active site is saturated at the higher substrate concentrations.

**Figure 6 | Comparing our Kemp eliminases with previous rational designs.** (a) The catalytic efficiencies ($k_{\text{cat}}/K_M$ values) of the single (W229D) and double (W229D/F290W) mutant variants of the ancestral β-lactamases studied here are shown in red. The red square data point represents the high pH value for GNCA4-W229D/F290W (Fig. 4a). The values for the ‘previous minimalist designs’ are taken from Korendovych et al. (Fig. 4a) (large orange data point) and Merski and Shoichet (small orange data points). In both cases, the design is based on a single mutation, but, in the latter study, additional mutations had to be introduced for mainly stabilization purposes (the values shown correspond to the variants displayed in Fig. 3 of Merski and Shoichet). The values obtained using a design approach that involved Rosetta are taken from Röthlisberger et al. Only the values for the eight designs that led to substantial Kemp eliminase activity are shown (59 designs were actually tested by Röthlisberger et al.). Iterative design values are taken from Privett et al. Here, ‘iterative’ means that an original design with very low activity was improved on the basis of 3D-structural information and molecular dynamics simulations to achieve substantial levels of catalysis. The efficiency of the best Kemp eliminase reported to date is also shown here for comparison. (b) Comparison of the Kemp eliminase turnover numbers ($k_{\text{cat}}$ values) obtained in this work with those reported for previous rational designs. The meaning of the symbols is the same as in a. The values shown as open symbols for the W229D variants of the ancestral backgrounds are actually lower limit estimates calculated as described in Supplementary Fig. S. We also show here the reaction rates corresponding to the best artificial Kemp eliminase reported to date and to the uncatalysed reaction for comparison.
acetate at the new active site, k is saturated.

Fig. 6). These are remarkable results, in particular, since, we are dealing here with a new active site generated on the basis of a minimalist approach that did not target esterase activity. Furthermore, the turnover number at the antibiotic degradation site. The esterase catalytic efficiency at the natural active site is, however, found to be ~20-fold smaller than that at the new active site.

at the higher substrate concentrations at which the de novo active site is saturated.

The catalytic efficiency for the hydrolysis of p-nitrophenyl acetate at the new active site, kcat/KM = 11.7 ± 1.1 M⁻¹s⁻¹, is similar to the best values reported in the literature for rationally designed esterases. Furthermore, the turnover number at the new active site, kcat = 7.6 × 10⁻³ ± 2 × 10⁻⁴ s⁻¹, indicates a three orders of magnitude enhancement over the rate of the non-enzymatic reaction (kuncat = 2.5 × 10⁻⁶ s⁻¹; Supplementary Fig. 6). These are remarkable results, in particular, since, we are dealing here with a new active site generated on the basis of a minimalist approach that did not target esterase activity.

Computational modelling of engineered ancestral enzymes. To further explore and confirm the role of conformational flexibility in the emergence of new enzyme functions, we have complemented our experimental work with molecular dynamics (MD) simulations of the wild-type and mutant forms of the modern TEM-1 and Bacillus licheniformis (BL) β-lactamases, as well as the ancestral ENCA, GNCA MP, GNCA and PNCA β-lactamases (Fig. 1). We chose these particular proteins both due to the availability of X-ray structures (see the Methods section) and because they span the range of Kemp elimination activities obtained in this work upon new active site generation. In particular, the modern TEM-1 and BL β-lactamases, as well as the comparatively recent ENCA β-lactamase (Fig. 1), display negligible activity upon the W229 mutation, while this mutation confers substantial Kemp elimination activity to the PNCA, GNCA MP and GNCA backgrounds following the order PNCA < GNCA MP < GNCA. Furthermore, the double W229/F290W variant of the GNCA scaffold is the most active Kemp eliminase reported in this work with a rate enhancement of about seven orders of magnitude over the rate of the uncatalysed reaction. As described below, our MD simulations support that conformational flexibility contributes substantially to these activity trends.

We have used the root mean square fluctuation (RMSF) of each amino acid in our simulations as a measure of the overall flexibility of the system. We examined first the GNCA MP and TEM-1 β-lactamases, that is, the two proteins for which NMR relaxation data are available (Fig. 3) from this and previous work, respectively. As explained in the second section of the Results, the region of the ancestral scaffold structure targeted for new active site generation is characterized by a large number of residues with conformational exchange contributions to the relaxation rates. Clearly, complete agreement between the results of the MD simulations and the conformational exchange contributions to the NMR relaxation rates is not to be...
expected, mainly because of the different timescales involved. That is, inclusion of exchange terms to explain relaxation rates reveals dynamic processes in the micro to milliseconds range\(^\text{33}\), while shorter timescales are typically probed by MD simulations\(^\text{33}\). Nevertheless, we find a clear correspondence between the calculated \(C_\text{r}\) RMSF values and the number of residues with conformational exchange contribution to NMR relaxation at the region of the new active site (Supplementary Fig. 8). It is therefore likely that the approximately microsecond conformational transitions at the region of the new active site in the ancestral scaffold are reflected in our 200 ns (\(\approx 0.2 \mu\)s) MD simulations, which actually approach the lower end of the conformational exchange timescale range.

The capability of our MD simulations to capture flexibility features that are relevant for new active site generation is further supported by the general congruence with the catalytic properties of several modern and ancestral scaffolds. In the case of the wild-type enzymes, a clear increase can be seen in the flexibility of residues 252 through to the h11 helix upon moving from TEM-1 and BL to the GNCA\textsubscript{MP} to GNCA4 \(\beta\)-lactamases, which are the residues that are primarily involved in forming the cavity creating the \textit{de novo} active site (Fig. 8). The flexibility of this region appears indeed to follow the general activity trend described above. Also, the largest overall changes in flexibility can be seen in helices h1 and h11, as well as the loops covering the \textit{de novo} active site, in agreement with both the crystallographic data (Fig. 4) as well as with the NMR relaxation data shown in Fig. 3. The mobility of different regions of these enzymes has also been highlighted in Fig. 9. The function-generating mutation W229D appears to increase the flexibility, which is the likely outcome of the introduction of a charged residue in a hydrophobic environment\(^\text{6}\). Still, the effect is much more pronounced in the ancestral \(\beta\)-lactamases, while the modern \(\beta\)-lactamases show only a comparatively small increase in flexibility (Fig. 8).

We subsequently performed MD simulations in the presence of the transition-state analogue 5(6)-nitrobenzotriazole (TSA) to explore the impact of transition-state binding on the conformational dynamics, as well as how well each system could accommodate the transition state in its new active site. The observed TSA binding features do correlate with the activity trends and with the flexibility trends defined by the RMSF data. The W229D variants of the ancestral scaffolds that give rise to substantial Kemp elimination activity (PNCA, GNCA\textsubscript{MP} and GNCA4) can accept the TSA in the MD simulations, although the TSA dissociates from the less active W229D variant of the PNCA scaffold after about 50 ns (Fig. 10). By contrast, efficient TSA binding is not observed with the W229D variants that show very low Kemp elimination activity. Specifically, we could not obtain a stable TSA complex with the W229D variant of ENCA \(\beta\)-lactamate, while the TSA flies out of the cavity in the W229D variants of the modern TEM-1 and BL \(\beta\)-lactamases within about 20 ns (Fig. 10). We have also observed the same pattern (low flexibility in the region of the \textit{de novo} active site and inability to retain the TSA bound) for the \(\beta\)-lactamases from \textit{Enterobacter cloacae} (Nmc-A) and \textit{Proteus vulgaris} (Bla-B) (Supplementary Fig. 9). This pattern, therefore, is likely to be a general feature of modern \(\beta\)-lactamases from Enterobacteria (Fig. 1).

Finally, there are two apparent exceptions to the flexibility/activity trend discussed above. First, ENCA \(\beta\)-lactamase develops substantial flexibility in the new active site region upon the W229D mutation, even though this protein shows negligible Kemp elimination activity. However, D229 is partially deprotonated at neutral pH and flexibility upon introduction of a charged group in a hydrophobic environment reflects, to some extent, the conformational fluctuations required to allow some water penetration and the consequent stabilization of the partially buried charge\(^\text{6}\). Such a specific kind of flexibility might not be relevant to the understanding of transition-state binding. Indeed, there should not be significant water penetration when the transition state is bound to the new active site, and the negative charge of the catalytic aspartate is actually stabilized by the interaction with the proton being abstracted. Second, the GNCA4 scaffold appears to be more rigid (in the new active site region) than the GNCA\textsubscript{MP} scaffold, despite the fact that GNCA4 leads to higher levels of Kemp elimination activity upon the W229D mutation. However, 3D-structure determination shows that the GNCA4 is actually more preorganized for transition-state binding than the GNCA\textsubscript{MP} scaffold (Fig. 4). This can again ultimately be viewed as a reflection of flexibility at the GNCA

---

**Figure 8 | \(C_\text{r}\)-atom root mean square fluctuations of all simulated variants.** Shown here are the root mean square fluctuations (RMSF, Å) of all \(C_\text{r}\) atoms during our simulations of (a) wild-type and (b) variant forms of the TEM-1 (blue), \textit{Bacillus licheniformis} (purple), GNCA\textsubscript{MP} (red), GNCA4 (orange), ENCA (dark green) and PNCA (light green) \(\beta\)-lactamases. In the case of the variant forms, the simulations were performed both in the absence of the transition-state analogue (TSA) or in complex with the TSA, 5(6)-nitrobenzotriazole. The simulations shown in (b) were all performed in the absence of the TSA, and a corresponding comparison of the \(C_\text{r}\) RMSF values of the ancestral enzymes both with and without the TSA present are shown in Supplementary Fig. 29. All values shown here are averages over the last 60 ns of three independent trajectories (180 ns total simulation time), which were obtained as described in the Methods section.
fluctuations (RMSF) of their Ca and PNCAb described in the Methods section. Molecular dynamics simulations, obtained from clustering analysis as shown here are the most representative structures obtained from our Bacillus licheniformis maps on to their tertiary structures. Note here that the mobilities of the different enzymes relative to each other, and how this in Å. This figure is presented to allow for a visual comparison of the relative mobility of the different enzymes relative to each other, and how this maps on to their tertiary structures. Note here that the Bacillus licheniformis β-lactamase has unstructured C- and N-terminal segments. The structures shown here are the most representative structures obtained from our molecular dynamics simulations, obtained from clustering analysis as described in the Methods section.

node, as mutations removed from the active site can easily shift the conformational equilibria at this node. Clearly, the specific mutations present in the GNCA4 scaffold have led to the population of a conformation in the W229D variant with a cavity where the substrate can easily form a reactive conformation, and with better architecture for effective transition-state stabilization.

Discussion

Modern proteins can perform an enormous diversity of molecular tasks, often with high efficiency and specificity. Most of these modern functions evolved from previously existing functionalities. Yet, it is inescapable to assume that the emergence of completely new functions has also occurred, at least at some early stages in protein evolution. For instance, it has been recently estimated from the number of unique enzymes and the number of domain superfamilies that at least 87% of all enzyme functions have evolved from another function or from ancestors with generic functionalities. This would leave a fraction of 13% of enzyme functions as plausible candidates for having emerged through the generation of new active sites. While the 13% figure may be an overestimation, as some cases of homology between superfamilies may be missed by sequence comparison, it does however suggest that the generation of new active sites may not actually be an exceedingly rare event and thus should garner more serious consideration. This is supported by reports of alternate-site promiscuous enzymes that display catalysis of a secondary reaction at a site other than the active site of the natural catalytic process. Furthermore, most fundamental biochemical processes are extremely slow in the absence of enzymes and seemingly specialized enzymes are likely to have catalysed many different reactions already in the last universal common ancestor. It may be reasonably inferred that efficient mechanisms for the emergence and subsequent evolution of new enzyme functionalities must exist or, at least, that they must have existed at the primordial protein stage. Little is known, however about these mechanisms. Recent work has shown that introducing a catalytic group in the hydrophobic cavity of the C-terminal domain of calmodulin can generate significant levels of catalysis for simple reactions. The simple mechanism for the emergence of new enzyme functions demonstrated here does not require the recruitment of a preexisting hydrophobic cavity. A hydrophobic-to-ionizable amino acid replacement generates a buried (or partially buried) residue with the perturbed properties that are useful in catalysis and conformational flexibility assists the generation of a completely new active site by facilitating substrate and transition-state binding. We have shown that, when using resurrected Precambrian β-lactamas as scaffolds for protein engineering, a minimalist design approach based on this mechanism leads to levels of catalysis for the Kemp elimination reaction of up to approximately seven orders of magnitude above the rate of the uncatalysed reaction, as well as to significant ester hydrolysis activity. The role played by conformational flexibility in the generation of the new function is clearly apparent in the hydrophobic cavity. The process of inferring ancestral sequences inherently generates uncertainty. It is therefore customary in the field to test phenotypic robustness by studying not only the protein encoded by the most probabilistic sequence at a given node, but also the proteins encoded by alternative reconstructions derived from random sampling from the posterior probability distribution at the node. We have previously used this approach to demonstrate phenotypic robustness for the antibiotic degradation activity at the natural active site of the GNCA β-lactamase. Remarkably, however, the new active site behaves differently and engineered β-lactamas encoded by several alternative reconstructions at the GNCA node (common ancestor of the...
Gram-negative bacteria) display Kemp elimination activities that span a ~30-fold range. The several GNCA sequences studied differ among themselves at 8–20 positions, which are, in essentially all cases, distal from the neighbourhood of the Kemp elimination active site (Supplementary Table 3). Therefore, the ~30-fold range of catalytic efficiency observed suggests long-range communication between residues, a feature that is often linked to conformational dynamics. It emerges that function at new active sites may be more sensitive to long-range effects than function at ‘old’ highly evolved active sites. In any case, the existence of such long-range effects makes it difficult to make a meaningful comparison between the ancestral and the modern proteins, as our attempts to generate a substantial Kemp elimination activity in the modern β-lactamases by removing potential steric interferences to transition-state binding were unsuccessful (Supplementary Note). Therefore, the distinct ancestral versus modern pattern of new function generation found appears to provide direct evidence in support of recent proposals about the potential of resurrected ancestral proteins as scaffolds for protein engineering.

**Methods**

**Protein expression and purification.** Purification of the different β-lactamases studied in this work followed the procedures we have previously described with minor modifications. Briefly, the genes were cloned into a pET24 vector with kanamycin resistance and transformed into *Escherichia coli* BL21(DE3) cells. Ni-NTA affinity chromatography was used to purify the His-tagged proteins used in most of the experiments reported in this work. We checked that the presence of a long-range mutational effects may have helped avoid stasis between the improvements in the catalytic machinery caused by rare mutations at the active site.

Finally, it is remarkable that the minimalist approach that generates substantial levels of a new function in most of the ancestral β-lactamases tested here only leads to catalysis levels barely distinguishable from the background when performed on 10 modern β-lactamases (Supplementary Fig. 1). This different behaviour is plausibly linked to differences in conformational flexibility between the ancestral and the modern proteins, as our attempts to generate a substantial Kemp elimination activity in the modern β-lactamases by removing potential steric interferences to transition-state binding were unsuccessful (Supplementary Note). Therefore, the distinct ancestral versus modern pattern of new function generation found appears to provide direct evidence in support of recent proposals about the potential of resurrected ancestral proteins as scaffolds for protein engineering.
His-tag does not substantially impair catalysis (Supplementary Table 6 and NATURE COMMUNICATIONS | DOI: 10.1038/ncomms16113 ARTICLE calculated using an extinction coefficient of 15,800 M
between the rates determined in these buffer systems at overlapping pH values.

Activity assays. Kemp elimination activity assays were performed at 25 °C in HEPES 10 mM, 10 mM sodium phosphate or 10 mM borate (in all cases with (GNCAMP-W229D), 2 M NH₄SO₄ 0.1 M Tris-HCl pH 8.0 (GNCA MP) or a PEG analogue-liganded forms was unsuccessful. Therefore, the liganded forms were capillary and equilibrated with the mother liquid supplemented with 15% (v/v) of those of specific amino acid type discrimination. The only unas-
lar signals were made using standard procedures with the following experiments:

Protein crystallization and structure determination. Crystals of GNCA MP (WT and W229D mutant, 25 mg ml⁻¹) β-lactamas were grown in capillaries by the counter-diffusion technique with 5 M sodium formate, 0.1 M sodium acetate pH 4.0 (GNCA MP-W229D) 5% (v/v) glycerol, 0.1% (v/v) Tri-HCl pH 8.0 (GNCA MP-PEG (polyethylene glycol) mixture (PEG 400/4000/8000; 20, 10 and 15%, respectively), Tris-HCl 0.1% pH 7.0 (GNCA MP and GNCA MP-W229D/F290W) as precipitating agents. In the case of the unliganded forms, crystals were extracted from the capillary and equilibrated with the mother liquid supplemented with 15% (v/v) glycerol before flash-cooling them in liquid nitrogen. Co-crystallization of the TS-analogue-ligated forms was unsuccessful. Therefore, the liganded forms were obtained by immersion of a portion of capillary containing crystals of GNCA MP-W229D or GNCA MP-W229D/F290W into a solution containing the original precipitant composition but at pH 9.0, supplemented with 1 M of the transition-state analogue (5(6)-nitrobenzotriazole and 15% (v/v) glycerol. The crystals were equilibrated for 2 weeks being extracted flash-cooled. Data were collected at beam lines ID30A, ID29 and ID23-1 (ESRF) and XALOC (ALBA). The procedures used for 3D-structure determination, as well as the data collection and refinement statistics, are summarized in Supplementary Table 8 see also Supplementary Fig.14–19 for stereo plots of relevant portions of the electron density maps and Supplementary Fig. 20 for a structural comparison between the modern and ancestral β-lactamas in the region of the de novo active site. The resulting coordinates and structure factors have been deposited at the PDB with the accession codes 5FQJ, 5FQQ, 4UHU, 5FQK and 5FQM.

NMR spectroscopy. We provide here a brief summary of the NMR studies on the GNCA MP β-lactamase. For a more detailed account, Supplemental Methods, Supplementary Fig. 32 and Supplementary Tables 9–11, we also performed simulations on the BL-W229D, NMC-A, W229D-M182T-F290W and TEM-1-M182T-W229D variants; however, as crystal structures of three enzymes were unavailable, we manually inserted these substitutions into the wild-type enzyme using PyMOL's Mutagenesis assignments were made using standard procedures with the following experiments: 2D 1H-15N HSQC and 3D HNCO, HN(CA)CO, HN(CO)CA, HN(CA), HN( CA), CBCA(CO)NH and HN(CA) data. Obtained with these experiments were complemented with those of specific amino acid type discrimination. The only unassign-
with 100 mM NaCl), depending on the pH range. We found no significant difference between the rates determined in these buffer systems at overlapping pH values.

Product formation was followed spectrophotometrically at 380 nm and rates were calculated using an extinction coefficient of 17,400 M
on the basis of the free enzyme is probed by the catalytic efficiency versus pH profile, while the pK in the Michaelis–Menten complex is probed by the catalytic efficiency versus pH profile.

Degradation of the antibiotic nitrocefin was assayed at 25 °C in HEPES buffered 10 mM, 100 mM NaCl, pH 7.0 and followed spectrophotometrically at 486 nm. A change in the extinction coefficient of 17,400 M
were used to calculate degradation rates from the time dependence of the absorbance. Degradation of benzylpenicillin and cefotaxime was assayed as we have previously described in detail11. Inhibition of the antibiotic degradation activity by clavulanic acid, sulbactam and tazobactam was tested by incubation with 1 mM inhibitor at 4 °C. The inhibitor dependence of the catalysis was fitted using the following equation:

\[
\text{rate} = \frac{A \cdot 10^{(p_I - p_K)}}{1 + 10^{(p_I - p_K)}}
\]

where 'rate' stands for the catalytic efficiency or for the rate at a given substrate concentration. A and pK are fitting parameters. Equation (1) assumes that the pH dependence of catalysis is determined by a single pK. A simple analysis based on the Michaelis–Menten mechanism shows that the pK in the free enzyme is probed by the catalytic efficiency versus pH profile, while the pK in the Michaelis–Menten complex is probed by the catalytic efficiency versus pH profile.

We did not observe, however, the burst phase linked to the formation of an acyl intermediate.

The procedures used for 3D-structure determination, as well as the data collection and refinement statistics, are summarized in Supplementary Table 8. Both proteins appear very rigid on the ps-ns timescale with an order parameter of S² = 0.9 for all residues in both cases. There is, however, a clear difference on the timescale (us-ms) common to both biologically active proteins: allostery, allosteric regulation and catalytic cycle17. Specifically, the number of residues for which the relaxation rates cannot be explained without including a conformational exchange contribution (K_x) is 26 for GNCA MP-β-lactamase versus 12 for TEM-1-β-lactamase (Fig. 3).

Synthesis of 5(6)-nitrobenzotriazole. This compound was obtained following the reaction conditions reported by Wasik et al. Briefly, a solution of NaN₃ (230 mg, 3.3 mmol) in 1.5 ml of water was added to a mixture of 4-nitro-o-
phenylenediamine (306 mg, 2 mmol) in 4 ml of AcOH and 1.5 ml of water at 0°C. The solid was collected by filtration and washed with H₂O to afford 300 mg (91% yield) of pure light yellow solid that was dried under vacuum over P₂O₅, mp 213-214 °C (lit. 215-216 °C). 1D-NMR spectrum of the compound is given in Supplementary Fig. 23.

Spectrophotometric monitoring of ligand binding. Binding of the transition-state analogue to variant β-lactamas was followed on the basis of the inhibition of the Kemp elimination activity or esterase activity of this protein. Briefly, stock solutions of the transition-state analogue and indole in acetonitrile (approximate concentrations 0.1 M and 5 M, respectively) were prepared and different microliter amounts were added to 2 ml solutions of the β-lactamase. After 1 h incubation, activity was determined as described above. Longer incubation times led to essentially identically solutions. The final acetonitrile concentration in the assay solution was always below 2%.

Dissociation constants for inhibitors were determined from the fitting to the experimental data of an equation based on the Michaelis–Menten mechanism with reversible competitive inhibition:

\[
\text{rate} = \frac{A}{1 + S/K_s + I/K_i}
\]

where [S] and [I] are the substrate and inhibitor concentrations, A is a constant and K_i is the inhibitor dissociation constant. For W229D variants of the GNCA MP scaffold, plots of rate versus substrate concentration are linear (Supplementary Fig. 24), implying that S/K_i is much smaller than unity and can be neglected in the denominator of equation (2). For the W229D/F290W variants, the K_i in equation (2) was fixed in the value obtained from the Michaelis–Menten analysis of the rate versus substrate concentration profiles (Supplementary Fig. 25).
Wizard46. The specific GNCA variants were selected on the basis of their observed activities (Supplementary Table 6), and the variant simulations were performed both in the relaxed backbone- and side-chain state analogues (Supplementary Fig. S5), nitrobenzoxadiazole (PDB IDs for the β-lactamase-TSA complexes: SFQJ for GNCAPM-W229D, 5FQK for GNC4A-W229D/F290W), as well as in the TSA free form. For all other variants, the TSA was manually placed in the newly created cavity in agreement with structural alignment of the crystal structures from the corresponding TBSA containing structures. All MD simulations were performed with the GPU implementation of the AMBER16 simulation package47, and the ff14SB force field48,49. The Sander module was used for the initial minimization and equilibration runs, and PMEMD for the production runs in explicit solvent. To obtain ff14SB compatible parameters for the TSA, we parameterized this compound using ANTECHAMBER52 and Gaussian09 (ref. 53). Partial charges were obtained using the standard RESP fitting procedure54,55, and the ligand atoms were described as the GAFF force field48. All non-standard parameters generated by the GAFF force field are provided in the Supplementary Table 12.

All model systems were generated with LEaP. Each model system was placed in an octahedral TIP3P water box56,57, which allowed the box to extend to at least 11 Å from the solute in each direction. The protonation state of all ionizable residues was determined in this work have been deposited at the PDB with the accession codes 6N1H–W229D/F290W), as well as in the TSA free form. For all other variants, the TSA was manually placed in the newly created cavity in agreement with structural alignment of the crystal structures from the corresponding TBSA containing structures. All MD simulations were performed with the GPU implementation of the AMBER16 simulation package47, and the ff14SB force field48,49. The Sander module was used for the initial minimization and equilibration runs, and PMEMD for the production runs in explicit solvent. To obtain ff14SB compatible parameters for the TSA, we parameterized this compound using ANTECHAMBER52 and Gaussian09 (ref. 53). Partial charges were obtained using the standard RESP fitting procedure54,55, and the ligand atoms were described as the GAFF force field48. All non-standard parameters generated by the GAFF force field are provided in the Supplementary Table 12.

Walter et al. 1. Korendovych, I. V. & DeGrado, W. F. Catalytic efficiency of designed catalytic proteins. Curr. Opin. Struct. Biol. 27, 113–121 (2014).

2. Wolfenden, R. & Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res. 34, 938–945 (2001).

3. Roncali, J., Campo, A., Kuntz, I. & Gavara, J. A. Thermostable β-lactamases: past, present and future. Curr. Opin. Chem. Biol. 108, 6823–6827 (2011).

4. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503, 418–421 (2013).

5. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. J. Am. Chem. Soc. 137, 14905–14911 (2015).

6. Shape evolutionary and analysis. Curr. Opin. Struct. Biol. 19, 120–127 (2009).

Data availability. X-ray crystallographic data for the several 3D-structures determined in this work have been deposited at the PDB with the accession codes 5FQJ, 5FQQ, 4HUH, 5FQJ, 5PQK and 5FQM. Force field files for the transition-state analogue, as well as sample simulation input files and PDBs have been uploaded to Dryad, and can be accessed using the following DOI: 10.5061/dryad.53629. Additional data supporting the findings of this study are available within the article and its Supplementary Information files, and from the corresponding authors upon reasonable request.

References

1. Korendovych, I. V. & DeGrado, W. F. Catalytic efficiency of designed catalytic proteins. Curr. Opin. Struct. Biol. 27, 113–121 (2014).

2. Wolfenden, R. & Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res. 34, 938–945 (2001).

3. Roncali, J., Campo, A., Kuntz, I. & Gavara, J. A. Thermostable β-lactamases: past, present and future. Curr. Opin. Chem. Biol. 108, 6823–6827 (2011).

4. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503, 418–421 (2013).

5. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. J. Am. Chem. Soc. 137, 14905–14911 (2015).

6. Shape evolutionary and analysis. Curr. Opin. Struct. Biol. 19, 120–127 (2009).

Data availability. X-ray crystallographic data for the several 3D-structures determined in this work have been deposited at the PDB with the accession codes 5FQJ, 5FQQ, 4HUH, 5FQJ, 5PQK and 5FQM. Force field files for the transition-state analogue, as well as sample simulation input files and PDBs have been uploaded to Dryad, and can be accessed using the following DOI: 10.5061/dryad.53629. Additional data supporting the findings of this study are available within the article and its Supplementary Information files, and from the corresponding authors upon reasonable request.

References

1. Korendovych, I. V. & DeGrado, W. F. Catalytic efficiency of designed catalytic proteins. Curr. Opin. Struct. Biol. 27, 113–121 (2014).

2. Wolfenden, R. & Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res. 34, 938–945 (2001).

3. Roncali, J., Campo, A., Kuntz, I. & Gavara, J. A. Thermostable β-lactamases: past, present and future. Curr. Opin. Chem. Biol. 108, 6823–6827 (2011).

4. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503, 418–421 (2013).

5. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. J. Am. Chem. Soc. 137, 14905–14911 (2015).

6. Shape evolutionary and analysis. Curr. Opin. Struct. Biol. 19, 120–127 (2009).
Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M. & Orengo, C. A. NATURE COMMUNICATIONS | DOI: 10.1038/ncomms16113 | www.nature.com/naturecommunications

58. Sœndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H.

55. Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollman, P. A. Application of RESP et al. Gaussian 09; Revision. D.01

51. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C.

34. Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M. & Orengo, C. A. Large-scale analysis exploring evolution of catalytic machinery and mechanisms in protein superfamilies. J. Mol. Biol. 428, 253–267 (2016).

52. Palmer, III A. G. pdivineta (Columbia University) Available from: http://www.palmer.lsc.columbia.edu/software/diffusion.html.

59. Ragland, D. A. et al. Drug resistance conferred by mutations outside the active site through alerted changes in the dynamic and structural ensemble of HIV-1 protease. J. Am. Chem. Soc. 136, 11956–11963 (2014).

40. Devamani, T. et al. Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. J. Am. Chem. Soc. 138, 1046–1056 (2016).

42. Cole, R. & Loria, J. P. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. J. Biomol. NMR 26, 203–213 (2003).

46. The PyMOL Molecular Graphics System, Version 1.7.1 Schroeder, LLC.

37. Laurino, P.

38. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of mechanisms in enzyme superfamilies. J. Am. Chem. Soc. 138, 8577–8593 (1996).

39. Ragland, D. A. et al. Drug resistance conferred by mutations outside the active site through alerted changes in the dynamic and structural ensemble of HIV-1 protease. J. Am. Chem. Soc. 136, 11956–11963 (2014).

41. Otaíora, F., Gavira, J. A., Ng, J. D. & García-Ruiz, J. M. Counterdiffusion methods applied to protein crystallization. Prog. Biophys. Mol. Biol. 101, 26–37 (2009).

43. Cole, R. & Loria, J. P. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. J. Biomol. NMR 26, 203–213 (2003).

44. Lipkowitz & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. 104, 4546–4559 (1982).

45. Saoter, A., Kuzmanic, A., Pannu, N. S. & Zagrovic, B. X-ray refinement significantly improves the accuracy of protein side chain and backbone parameters from (P99SB. J. Chem. Theory Comput. 11, 3689–3713 (2015)

47. Case, D. A.

56. Söndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H.

57. Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollman, P. A. Application of RESP et al. Gaussian 09; Revision. D.01

50. Götz, A. W. et al. Routine microscopic molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. 8, 1542–1555 (2012).

53. Frisch, M. J. et al. Gaussian 09; Revision. D.01 (Gaussian Inc, 2009).

54. Cooper, W. D., Cieplak, P., Bayley, C. I. & Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies and free energies of solvation. J. Am. Chem. Soc. 115, 9620–9631 (1993).

55. Cornwell, W. D., Cieplak, P., Bayley, C. I. & Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies and free energies of solvation. J. Am. Chem. Soc. 115, 9620–9631 (1993).

60. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the molecular dynamics switch to an external bath. J. Chem. Phys. 79, 926 (1983).

61. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microscopic molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. J. Chem. Theory Comput. 9, 3878–3888 (2013).

62. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanics calculations. J. Mol. Graph. Mod. 25, 247–260 (2006).

63. Frisch, M. J. et al. Gaussian 09; Revision. D.01 (Gaussian Inc, 2009).

64. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. 9, 3084–3095 (2013).

65. Halle, B. Flexibility and packing in proteins. Proc. Natl Acad. Sci. USA 99, 1274–1279 (2002).

66. DePristo, M. A., Bakker, P. I. W. & Blundell, T. L. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 12, 831–838 (2004).

67. Kuzmanic, A., Pannu, N. S. & Zagrovic, B. X-ray refinement significantly understimates the level of microscopic heterogeneity in biomolecular crystals. Nat. Commun. 5, 3220 (2014).

68. Rueda, M. et al. A consensus view of protein dynamics. Proc. Natl Acad. Sci. USA 104, 796–801 (2007).

Acknowledgements

This work was supported by Feder Funds, Grants from the Spanish Ministry of Economy and Competitiveness BIO2015-64642-R (J.M.S.-R.), CSD2009-00088 (J.M.S.-R.), CTQ2011-29299-C02-01 (F.S.-G.), CTQ2011-22514 (M.B.), BIO2016-74875-P (J.A.G.), ‘Factoría Española de Crystalización’, Consolider-Ingenio 2010 (J.A.G.) and CEI BioTic V19-2015 (V.A.R.), a Wallenberg Academy Fellowship (S.C.L.K.) and DuPont Young Professor Award (E.A.G.) and Grants NX131A08G and NX131A11G (E.A.G.) from NASA Exobiology. The European Research Council has provided financial support under the European Community’s Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 306474. We acknowledge the ESRF and ALBA for provision of synchrotron radiation time at beam lines ID29, ID23-1 and ID30-A-1, and Xaloc, respectively, and the staff for their helpful support. Finally, we are grateful to the Swedish National Infrastructure for Computing (SNIC, 2015/16–12) for their generous provision of computational resources.

Author contributions

V.A.R. obtained the protein variants, and designed, performed and analysed the experiments addressed at determining their natural and non-natural activities. S.M.-R. crystallized protein variants and determined their X-ray structure under the supervision of I.A.G., who also provided essential input regarding the interpretation of the 3D-structures. A.M.C. and D.P.-U. performed and analysed NMR relaxation experiments under the supervision of M.B. who also provided essential input regarding the interpretation of the relaxation data. D.M.K. performed the MD simulations under the supervision of S.C.L.K., who also provided essential input regarding the properties of the transition-state analogue. E.A.G. provided methodological support for bioinformatics analyses, as well as essential input for the interpretation of the results obtained in an evolutionary context. J.M.S.-R. and V.A.R. designed the research, with the exception of the computational work that was designed by S.C.L.K., J.M.S.-R., V.A.R. and S.C.L.K. wrote the manuscript. All authors discussed the manuscript and suggested modifications and improvements.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

How to cite this article: Risso, V. A. et al. De novo active sites for resurrected Precambrian enzymes. Nat. Commun. 8, 11613 doi: 10.1038/ncomms16113 (2017).

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

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/

© The Author(s) 2017