ARTICLES
PUBLISHED ONLINE: 29 AUGUST 2016 | DOI: 10.1038/NCHEM.2596
nature chemistry

Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase

Richard Obexer[1†], Alexei Godina[1,2†], Xavier Garrabou[1], Peer R. E. Mittl[3], David Baker[4], Andrew D. Griffiths[2*] and Donald Hilvert[1*]

Designing catalysts that achieve the rates and selectivities of natural enzymes is a long-standing goal in protein chemistry. Here, we show that an ultrahigh-throughput droplet-based microfluidic screening platform can be used to improve a previously optimized artificial aldolase by an additional factor of 30 to give a >10^9 rate enhancement that rivals the efficiency of class I aldolases. The resulting enzyme catalyses a reversible aldol reaction with high stereoselectivity and tolerates a broad range of substrates. Biochemical and structural studies show that catalysis depends on a Lys-Tyr-Asn-Tyr tetrad that emerged adjacent to a computationally designed hydrophobic pocket during directed evolution. This constellation of residues is poised to activate the substrate by Schiff base formation, promote mechanistically important proton transfers and stabilize multiple transition states along a complex reaction coordinate. The emergence of such a sophisticated catalytic centre shows that there is nothing magical about the catalytic activities or mechanisms of naturally occurring enzymes, or the evolutionary process that gave rise to them.

Enzymes are the most proficient catalysts known. Their remarkable activities and specificities reflect sophisticated active site geometries in which arrays of functional residues, organized in extended networks and tuned by outer-shell interactions, mediate substrate recognition and transition-state stabilization. This inherent complexity makes the design of enzymes for new tasks extremely challenging.

Class I aldolases, a large family of enzymes that promote reversible cleavage of carbon–carbon bonds[1], are a case in point. They utilize a multistep mechanism that involves covalent Schiff base adducts between an active-site lysine and key intermediates along the reaction coordinate. Preorganized sets of polar residues activate the lysine and orchestrate substrate binding, proton shuffling and stabilization of the individual transition states. As making and breaking C–C bonds are fundamental transformations in chemistry and biology, aldolases have found wide application as biocatalysts for asymmetric syntheses of biologically important compounds[2–4]. Efficient designer enzymes with tailored substrate specificities could, however, extend their biotechnological utility substantially.

Simple organocatalytic amines[5], foldamers[6,7], catalytic antibodies[8,9] and computationally designed enzymes[10,11] have been developed as class I aldolase mimics. Although useful activities and selectivities have been realized, these catalysts are typically much less efficient than true enzymes, which reflects the relative simplicity of their designs. For some protein catalysts, it has been possible to boost low starting activities by directed evolution. Computationally designed retro-aldolases that cleave the fluorogenic substrate methodol (compound **1a** in Fig. 1), for example, have been improved more than 1,000-fold by repeated cycles of random mutagenesis and conventional plate screening[11,12]. The largest activity increases are generally observed in early optimization rounds, followed by diminishing returns as beneficial mutations become increasingly infrequent.

Droplet-based microfluidic systems[13] that exploit aqueous droplets as independent microreactors (Fig. 2) combine the flexibility of microtitre-plate assays and the ultrahigh throughput of fluorescence-activated cell sorting. Compartmentalization in droplets provides a simple means of linking genotype and phenotype[14]: encapsulating single cells that produce an enzyme of interest in picolitre-volume water droplets together with a fluorogenic substrate[15,16] and, if necessary, lysis agents[17] ensures that product formation occurs in the same compartment as the catalyst-encoding gene. The fluorescent product-containing droplets can then be sorted at frequencies up to 2,000 droplets per second using microfluidic fluorescence-activated droplet sorters (FADS)[15,16]. For comparison, screening ~2,000 aldolase variants for methodol cleavage in conventional plate assays took about two weeks[12]. Here we show that the increased throughput of FADS facilitates the discovery of aldolase catalysts that exhibit rates and selectivities characteristic of natural enzymes.

Results

Mutagenesis and FADS screening. As our starting point, we used the artificial retro-aldolase RA95.5-8, which catalyses the cleavage of methodol with a 3×10^7-fold rate acceleration[12]. This enzyme was obtained by extensive modification of the computationally designed retro-aldolase RA95.0 over 13 rounds of directed evolution. Structural characterization of key intermediates along the evolutionary trajectory highlighted the creation of an alternative substrate binding pocket and replacement of the original catalytic lysine, Lys210, with a new functional group, namely Lys83, as the source of its 4,000-fold higher activity[12]. Further optimization proved difficult, however, and the final three rounds of mutagenesis and microtitre-plate screening augmented specific activity only 2.5-fold.

To prepare RA95.5-8 libraries for FADS experiments, the encoding gene was diversified by a combination of error-prone PCR,

[1]Laboratory of Organic Chemistry, ETH Zürich, 8093 Zürich, Switzerland. [2]Laboratory of Biochemistry, École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI Paris), CNRS UMR 8231, 10, rue Vauquelin, 75231 Paris Cedex 05, France. [3]Department of Biochemistry, University of Zürich, 8057 Zürich, Switzerland. [4]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. †These authors contributed equally to the work. *e-mail: hilvert@org.chem.ethz.ch; andrew.griffiths@espci.fr

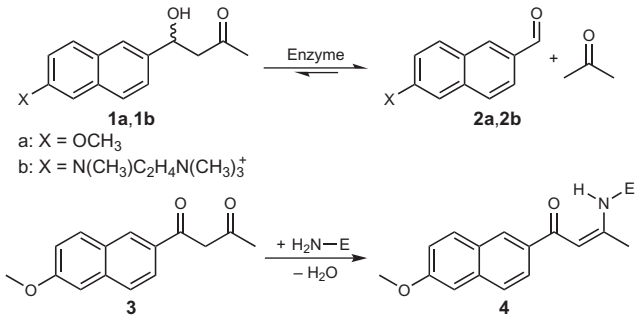


Figure 1 | Substrates and inhibitor for the computationally designed aldolase. The enzyme promotes the reversible interconversion of methodol derivatives (**1a** and **1b**) to the corresponding naphthaldehydes (**2a** and **2b**) and acetone. Diketone **3** inhibits catalysis by reacting covalently with the catalytic lysine (H_2N-E) to form the vinylogous amide **4**.

cassette mutagenesis and DNA shuffling. *Escherichia coli* cells expressing the gene libraries were compartmentalized directly with a fluorogenic substrate and lysis agents in surfactant-stabilized droplets dispersed in an inert fluorinated oil using a microfluidic system (Fig. 2). For screening, methodol (**1a**) was replaced with the positively charged aldol substrate **1b** (Fig. 1) to prevent diffusion of the fluorescent product **2b** between droplets.

To enhance the temporal resolution, we developed novel microfluidic devices that integrate droplet generation, incubation and sorting on a single chip (Fig. 2c). The incubation time was controlled by a delay line that could be varied in length depending on the activity of the starting catalyst[18]: selection stringency for ever-faster catalysts could thus be increased simply by reducing the incubation times from one hour to five minutes.

Using this set-up, we sorted ~10^6–10^7 library members per experiment. After sorting, the contents of the droplets were pooled, plasmid DNA from the enriched population amplified and individual clones analysed before proceeding to the next round. The best catalyst to emerge after six cycles of mutagenesis and screening (Supplementary Fig. 5), RA95.5-8F, was 30-fold more active than RA95.5-8 and contained 13 additional mutations (Supplementary Fig. 6a). These substitutions, which are scattered throughout the protein (Fig. 3a and Supplementary Fig. 6b), include a new active-site tyrosine (Phe180Tyr) and a leucine in place of the original catalytic lysine (Lys210Leu).

Kinetic characterization of the optimized variant. RA95.5-8F was subjected to steady-state kinetic analysis. The optimized enzyme catalyses the cleavage of (*R*)-methodol with a turnover number (k_{cat}) of 10.8 ± 0.6 s^{-1} and an apparent second-order rate constant, k_{cat}/K_M, of 33,800 ± 4,200 M^{-1} s^{-1} (Fig. 3b and Table 1). As seen previously for the directed evolution of a mechanistically simpler Kemp eliminase[19], efficiency was augmented not by decreasing the

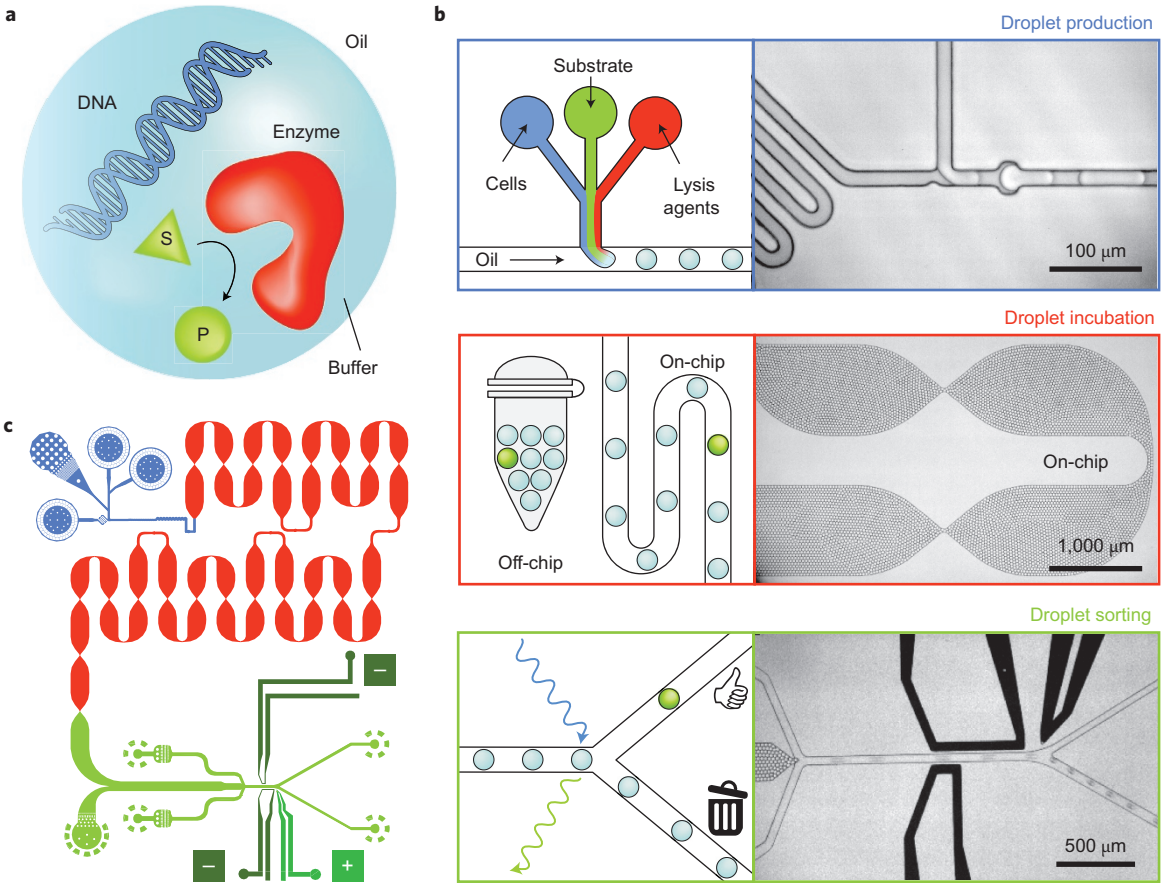


Figure 2 | Microfluidic platform for the directed evolution of aldolases. a, Aqueous droplets provide a physical link between DNA and proteins from lysed cells and product (P) generated from a fluorogenic substrate (S), and so allow genotype–phenotype coupling. **b**, The assay comprises three steps: (1) droplet formation to compartmentalize single cells with the fluorogenic substrate and lysis reagents; (2) incubation off-chip (for long incubations) or on-chip (for shorter incubations and enhanced temporal resolution); and (3) fluorescence-activated droplet sorting. **c**, Second-generation microfluidic chips integrate all three steps on a single device (blue, droplet nozzle; red, five-minute incubation line; green, sorting module). Supplementary Figs 1–4 give additional details regarding the microfluidic platform and chip designs.

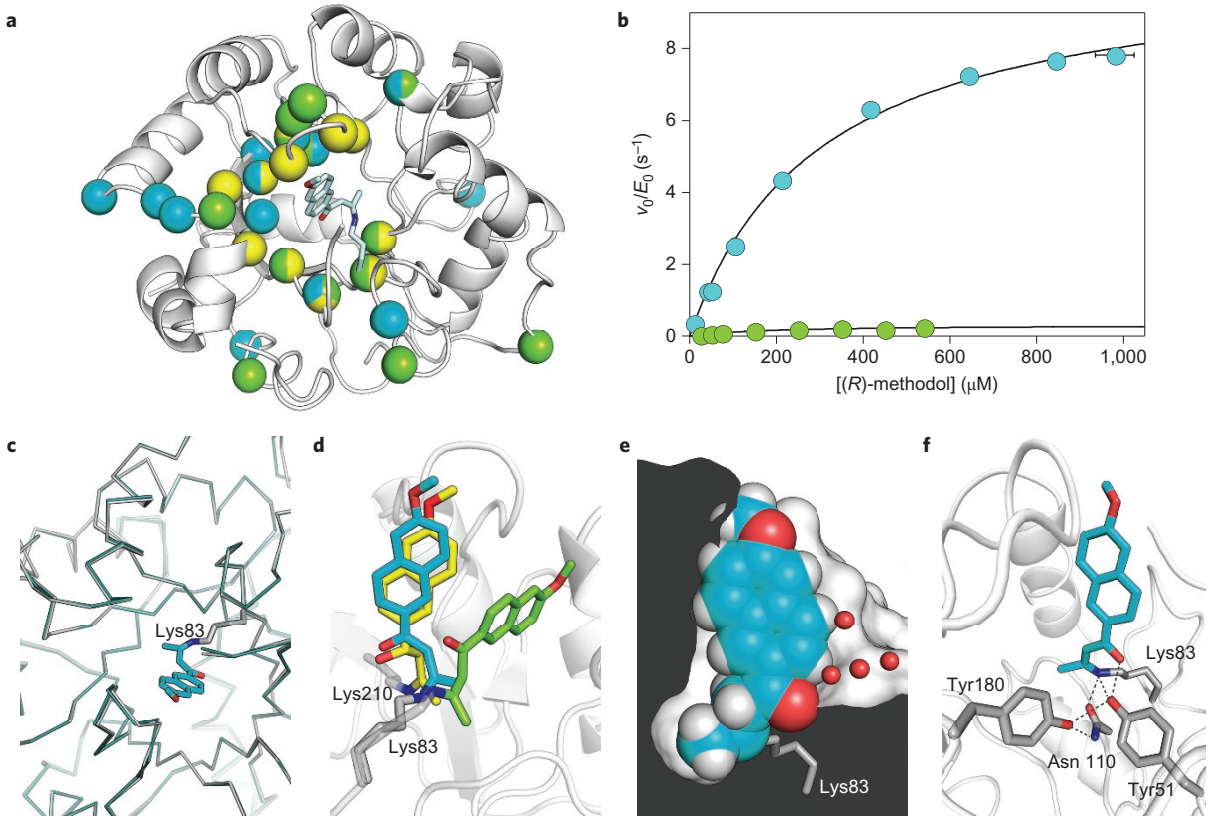


Figure 3 | Evolution and structural analysis of the optimized RA95.5-8F aldolase. a, The computationally designed RA95.0 retro-aldolase was generated by introducing 11 mutations (yellow spheres) into a $(\beta\alpha)_8$-barrel scaffold[11] and subsequently optimized by mutagenesis and plate screening (RA95.5-8, green spheres)[12] followed by microfluidic-assisted evolution (RA95.5-8F, cyan spheres). Multicolour spheres indicate residues that were mutated multiple times during optimization. **b**, Michaelis–Menten plots for RA95.5-8 (green) and RA95.5-8F (cyan). E_O, total enzyme concentration; v_O, initial reaction velocity. **c**, Alignment of the X-ray structures of RA95.5-8F with (grey) and without (teal) the 1,3-diketone ligand **3** bound at the active site. **d**, Overlay of the original theozyme (yellow) and inhibitor **3** bound at the active site of RA95.5-5 (green) and RA95.5-8F (cyan). RA95.5-5 is the structurally characterized precursor of RA95.5-8 (ref. 12). **e**, Cut-away view of the RA95.5-8F active site showing the snug fit of the ligand within the binding pocket. Small red spheres represent water molecules. **f**, Interactions (dashed lines) between the residues of the catalytic tetrad that emerged during directed evolution and the bound inhibitor.

Table 1 | Steady-state parameters of RA95 variants.

Enzyme*	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat}/k_{uncat}	$(k_{cat}/K_M)/k_{Lys}$	$S^{\dagger}$	pK_a
RA95.5-8F	10.8 ± 0.6	320 ± 36	34,000	1.7×10^9	1.5×10^{10}	480	6.2
RA95.5-8‡	0.36	230	1,600	5.5×10^7	7.0×10^8	14	7.6
RA95.5‡	0.0043	270	16	6.6×10^5	6.9×10^6	3	7.6
RA95.0‡	0.00005	300	0.17	4.8×10^3	2.3×10^4	0.4	8.1

Steady-state parameters were determined with (*R*)-methodol. All the measurements were performed at 29 °C in 25 mM HEPES, 100 mM NaCl, pH 7.5 containing 2.7% acetonitrile. Errors reflect the s.d. of 3–6 independent measurements. For comparison, $k_{uncat} = 6.5 \times 10^{-9}$ s^{-1} and $k_{Lys} = 2.3 \times 10^{-6}$ M^{-1} s^{-1} at 25 °C (ref. 27). *RA95.0 denotes the original computational design[11], RA95.5 the variant obtained after five rounds of active-site cassette mutagenesis[11] and RA95.5-8 the variant subsequently generated by eight rounds of error-prone PCR and DNA shuffling[12]. The highly active RA95.5-8F variant was identified in the current study. †Selectivity factor *S*, calculated as the ratio $S = (k_{cat}/K_M)_R/(k_{cat}/K_M)_S$. ‡Values from Giger *et al.*[12] RA95.5-8 is 2.5 times more active than the structurally characterized RA95.5-5 and contains three additional mutations.

Michaelis constant (K_M) but by increasing k_{cat}, which is considered the most demanding challenge in enzyme evolution[20].

Comparison with the uncatalysed retro-aldol reaction of methodol in buffer shows that RA95.5-8F achieves a 2×10^9-fold rate of acceleration (k_{cat}/k_{uncat}). It is also $>10^{10}$-fold more effective than simple amines like lysine [$(k_{cat}/K_M)/k_{Lys}$] and outperforms other artificial retro-aldolases that cleave **1a** by two-to-three orders of magnitude[8,9,11]. Moreover, the kinetic parameters determined for RA95.5-8F compare favourably with those for natural class I aldolases. For example, human fructose 1,6-diphosphate (FDP) aldolase catalyses a key step in glucose metabolism with $k_{cat} = 0.61$ s^{-1} and $k_{cat}/K_M = 265{,}000$ M^{-1} s^{-1} (ref. 21), whereas *E. coli* deoxyribose-5-phosphate aldolase (DERA), an enzyme widely used for synthetic applications[2], cleaves its native substrate with $k_{cat} = 68$ s^{-1} and $k_{cat}/K_M = 110{,}000$ M^{-1} s^{-1} (ref. 22).

In practical terms, the ability of RA95.5-8F to process up to 60,000 substrate molecules before being inactivated—a >40-fold improvement over RA95.5-8—is notable (Supplementary Fig. 7c), as are the evolved enzyme's thermostability ($T_m = 76$ °C) and tolerance to organic solvent (up to 14% acetone). Although screening was performed with racemic substrate, the preference for the (*R*)- over the (*S*)-enantiomer also increased in parallel with specific activity from 14:1 to 480:1 (Supplementary Fig. 7a,b). Kinetic optimization was evidently achieved by an increasingly precise recognition of the rate-limiting transition state for cleavage of the (*R*)-configured substrate.

High-resolution structural analysis. To understand the origins of the enzyme's high efficiency, we crystallized RA95.5-8F in the presence and absence of a mechanism-based inhibitor, 1,3-

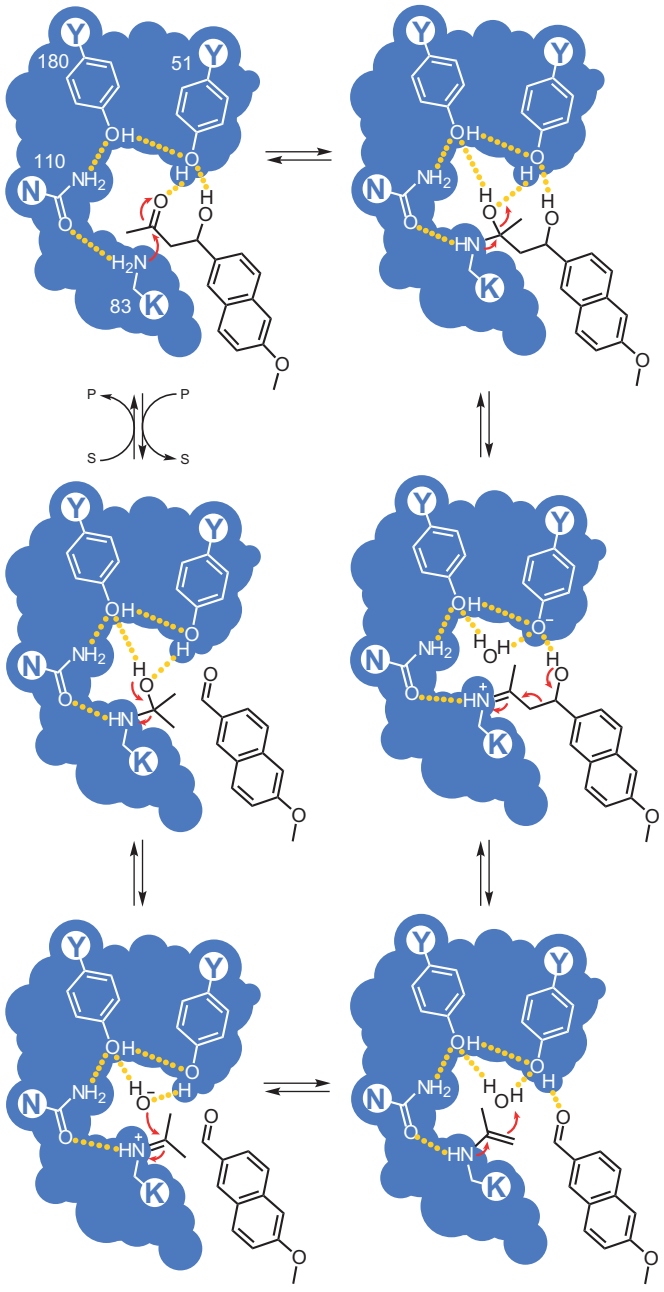


Figure 4 | Hypothetical mechanism for aldol cleavage by the evolved enzyme. The retro-aldol reaction occurs via a multistep sequence (starting in the upper left and proceeding clockwise) initiated by nucleophilic attack of Lys83 on the substrate. The supporting polar residues at the active site contribute to catalytic efficiency by activating the substrate, positioning the mechanistically important water molecule and promoting key proton transfers.

diketone **3** (Fig.1), and determined both structures to 1.10 Å resolution (Supplementary Fig. 8). The backbone and side-chain conformations of the free enzyme and inhibitor complex are very similar (root mean square deviation (r.m.s.d.) = 0.338 Å (Fig. 3c and Supplementary Fig. 9a,b)), indicating a high degree of preorganization, even at the active site.

In the liganded structure, compound **3** forms a covalent Schiff base adduct (**4**) with Lys83. Rather than adopt the planar configuration expected for a vinylogous amide, the ligand is slightly twisted (O–C–C–C dihedral angle, −29.1°), which allows the free carbonyl group to engage in a short hydrogen bond with Tyr51. More surprisingly, its naphthyl group does not dock in the hydrophobic pocket created during previous rounds of evolution[12]. Instead, it migrated back to the original computationally designed site (Fig. 3d,e). With one exception (Arg182Met), the 12 residues that line the designed apolar pocket were conserved over the entire evolutionary trajectory (Supplementary Fig. 10). Removal of the Lys210 side chain, which had extended into this site, in combination with an altered rotameric preference for Lys83 made reorientation of the ligand possible. It is striking how well the 'new' binding mode matches the initial computational model for the naphthyl ring (Fig. 3d). The agreement between the evolved structure and the original design (r.m.s.d. = 1.566 Å) is substantially better than that observed for the starting RA95.0 catalyst (r.m.s.d. = 2.659 Å), in which the naphthyl ring was rotated 114° along its long axis compared with the design model[12].

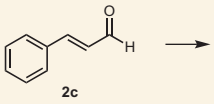
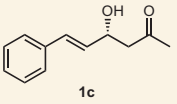
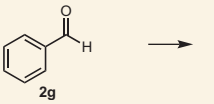
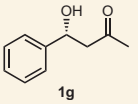
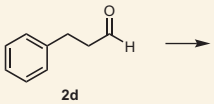
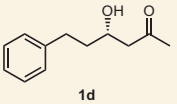
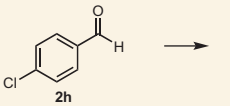
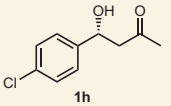
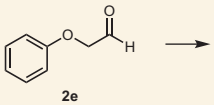
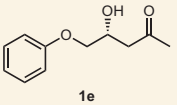
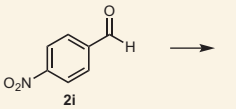
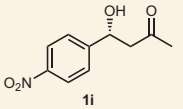
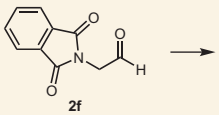
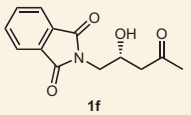
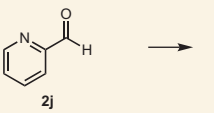
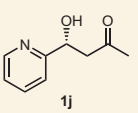
In contrast to the residues responsible for substrate recognition, the original catalytic apparatus, which consists of Lys210, Glu53 and an ordered water molecule, was completely erased over the course of evolution. In its place, a well-defined hydrogen-bonded network of four residues—Lys83, Tyr51, Asn110 and the newly installed Tyr180 (Fig. 3f)—emerged. Like the catalytic groups in natural $(\beta\alpha)_8$ barrel enzymes, the residues of this novel catalytic tetrad project into the active site from the ends of different β-strands[23,24]. The side chains of Tyr51 and Asn110 are oriented for effective catalysis through hydrogen-bonding interactions with Tyr180, second-shell residues and ordered water molecules (Supplementary Fig. 9c,d).

Multistep reaction mechanism. As shown schematically in Fig. 4, the catalytic tetrad is ideally poised to interact productively with bound substrate and adapt to geometric and electrostatic changes along the complex reaction coordinate. Thus, Lys83 can initiate the reaction by attacking the substrate carbonyl to form a tetrahedral carbinolamine intermediate that subsequently breaks down to give a protonated Schiff base. This iminium ion serves as an electron sink that facilitates cleavage of the adjacent C–C bond to generate an aldehyde and an enamine. Tyr51, deprotonated by the hydroxide ion released during the Schiff base formation, is well positioned to function as a catalytic base in this step (Fig. 3f). Subsequent protonation and hydrolysis of the enamine leads to the release of acetone and regeneration of the enzyme.

Several lines of evidence support this mechanism. First, Lys83 is essential for activity. Replacing it with methionine reduces the catalytic efficiency $>10^5$ fold (Supplementary Fig. 11a and Supplementary Table 1). Lys83 reactivity was also enhanced during microfluidic optimization, judging by the decrease in apparent pK_a from 7.6 to 6.2 (Supplementary Fig. 7d), a value similar to that observed for natural aldolases[22,25]. Second, residues Tyr51, Asn110 and Tyr180 form a hydrogen-bonded network capable of binding the hydroxyl group of the carbinolamine intermediate or, in later steps of the catalytic cycle, an ordered water molecule/hydroxide ion (Supplementary Figs 8a and 9c). Third, the use of Tyr51 to promote key proton transfers, which finds precedent in natural aldolases[22,26], rationalizes the enzyme's 480:1 selectivity for (*R*)- over (*S*)-methodol because this residue can make a productive interaction with the (*R*)- but not the (*S*)-configured alcohol. Although reverting Tyr51, Asn110 and Tyr180 individually results in modest reductions in k_{cat} (90-, 27- and fourfold, respectively (Supplementary Fig. 11b–d)), mutating Tyr51 together with Tyr180 decreases k_{cat} 1,700-fold (Supplementary Fig. 11e). Simultaneous reversion of all three residues causes an 8×10^5-fold drop in activity (Supplementary Fig. 11f). Such synergistic effects suggest functional coupling and are typical of highly evolved active sites.

Synthetic aldol reactions. Enantioselective catalysis of reactions that form C–C bonds has made natural aldolases useful tools in organic chemistry[2–4]. In contrast, computationally designed retro-

Table 2 | Substrate specificity of the evolved RA95.5-8F aldolase.

Aldehyde acceptor	Product	Conversion (%)	e.r.	Aldehyde acceptor	Product	Conversion (%)	e.r.
 2c	 1c	49	98.2:1.8	 2g	 1g	51	96:4
 2d	 1d	95	96:4	 2h	 1h	45	97.5:2.5
 2e	 1e	95	94:6	 2i	 1i	>95	84:16
 2f	 1f	88	98.8:1.2	 2j	 1j	95	83:17

Biocatalytic reactions of aldehyde acceptors (3–5 mM) with excess acetone (2–3 M) were conducted at 29 °C in 25 mM HEPES buffer, pH 7.5 containing 100 mM NaCl. The reactions were monitored by reverse-phase HPLC, and the 4*R*/4*S* enantiomeric ratio (e.r.) of the isolated aldol products was determined by chiral HPLC or supercritical fluid chromatography. The relative conversions reflect the equilibrium position of the respective synthetic/cleavage reactions.

aldolases, although effective in the cleavage direction, have had limited synthetic utility as yet, in part because they tend to be strongly inhibited by the aldehyde acceptor[27], which competes with the ketone donor for the active-site lysine and forms a non-productive Schiff base adduct. The high turnover numbers achieved by RA95.5-8F compared with those of its evolutionary precursors suggest that this problem was largely eliminated in the course of optimization. Indeed, the evolved enzyme is a practical biocatalyst for aldol synthesis (Supplementary Fig. 12). Using 2 M acetone to drive the thermodynamically unfavourable reaction ($K = 1.2 \pm 0.1$ M^{-1}), it transforms 2 mM 6-methoxy-2-naphthaldehyde into (*R*)-methodol in 67% conversion (71% is the theoretical maximum under these conditions). The product was isolated in 60% yield and greater than 98.4% e.e.

Like natural aldolases, which typically exhibit strict donor specificity but broad acceptor tolerance[4], RA95.5-8F catalyses aldol reactions of acetone with diversely functionalized aromatic and aliphatic aldehydes, which affords useful biocatalytic conversions and good-to-excellent enantioselectivities (Table 2). The generally high stereoselectivity of the synthetic aldol reactions suggests that most of the acceptors adopt a well-defined binding mode in the slot-like hydrophobic pocket that was designed to accommodate the naphthyl ring of methodol. Donor specificity is more restricted, in line with the buried and comparatively tight-fitting acetone-binding site. As this substrate preference is complementary to that of natural aldolases, which typically do not use acetone[28], it expands the set of aldol products that can be accessed through biocatalysis. Although polar donors, like dihydroxyacetone, are not accepted, cyclohexanone can replace acetone to some extent in the reaction with aliphatic aldehyde **2d** to generate diastereomeric adducts (84:14 d.r.). It should be possible to enhance this activity and further broaden donor scope through additional active-site engineering.

Discussion

Our findings demonstrate the feasibility of using a bottom-up approach to create an exceptionally active and synthetically useful artificial enzyme for a complex, multistep chemical transformation. In its structural and mechanistic sophistication, the evolved RA95.5-8F aldolase resembles natural enzymes like DERA[22] and FDP aldolase[25,26]. Although the specific sets of functional groups differ, these catalysts all exploit tightly coordinated networks of polar amino acids, located at comparable positions in the $(\beta\alpha)_8$ barrel scaffold[23,24], to activate their respective substrates and transform them into product.

From an evolutionary perspective, the stepwise emergence of the catalytic tetrad in the artificial aldolase shows how complex catalytic centres can be assembled from relatively simple starting points. Conversion of the original computational design into a highly active enzyme required screening of only ~10^8 protein variants—a small number compared with the vastness of sequence space, but greatly exceeding the capacity of standard microtitre plate assays.

Given the extensive active site modifications introduced by directed evolution (Fig. 3d), it is natural to ask to what extent design contributed to the final optimized catalyst. Our data show that the computational algorithms can effectively create shape-complementary hydrophobic binding pockets but are much less successful with extended networks of polar residues. Nevertheless, recent progress in producing atomically accurate serine-containing catalytic triads suggests that the latter problem may be ripe for reinvestigation[29]. Increasing the complexity of the initial design models by deliberately installing potential hydrogen-bond donors and acceptors in the vicinity of key catalytic residues could provide more active catalysts and, at the same time, make them more evolvable.

Our results further suggest that it may be possible to tackle binding and catalytic determinants separately. Segregation of these properties, as seen in RA95.5-8F, is common in other $(\beta\alpha)_8$ barrel enzymes and believed to contribute to the functional diversity of this protein family[30]. In practical terms, computation and/or directed evolution of the loops responsible for substrate recognition might be profitably exploited for the rapid diversification of RA95.5-8F with respect to substrate specificity and catalytic function[31,32]. In this way, families of artificial enzymes that retain the high activity of natural aldolases but exhibit expanded scope could be created.

Although computationally designed enzymes have not yet matched the efficiency of natural enzymes, directed evolution provides a practical method for adjusting and refining their properties[33]. In speeding

up searches of sequence space, FADS and other ultrahigh-throughput screening techniques[34–37] offer exciting opportunities to explore parallel evolutionary trajectories with different scaffolds, catalytic motifs and reaction chemistries. Such approaches are likely to become increasingly important as designers take on the challenge of creating catalysts for other mechanistically complex transformations.

Methods

Library generation. Gene libraries were created by random mutagenesis of the entire gene, cassette mutagenesis of the desired sites and DNA shuffling of verified mutants. Random mutations were introduced into the RA95.5-8 genes by error-prone PCR[38] using the GeneMorph II mutagenesis kit (Agilent) or the JBS dNTP-Mutagenesis kit (JenaBioscience). Cassette mutagenesis libraries were prepared by overlap-extension PCR according to a technique previously described[39]. DNA shuffling was performed according to the staggered extension process method[40] or by DNAse fragmentation and reassembly[41].

Microfluidic screening assay. Electrocompetent *E. coli* cells (XL1-Blue or BL21 Gold (DE3)) were transformed with the plasmid library and then cultured in supplemented M9 medium. After the induction of protein production with IPTG and further incubation, the transformants were harvested, washed with supplemented M9 medium and diluted to the desired density in the same medium but containing 30% Percoll and 2 U ml^{-1} DNAse I. Individual cells were encapsulated in 20 pl droplets along with substrate and lysis reagents (1 mg ml^{-1} lysozyme, 2 mg ml^{-1} polymyxin B, 10 mM EDTA and 5 mM EGTA). After subsequent incubation, either on-chip or off-chip, droplets were sorted according to their fluorescence. Microfluidic chip designs, flow rates and sorting parameters are specified in the Supplementry Information.

Crystallization and structure determination. RA95.5-8F and RA95.5-8F derivatized with the mechanism-based inhibitor **3** were crystallized by vapour diffusion in sitting drops. Crystals were cryoprotected in mother liquor complemented with 25% (v/v) ethylene glycol and flash frozen in liquid nitrogen. Diffraction data were collected at 100 K at the PXI beamline at the Swiss Light Source (Paul Scherrer Institute). Structures were solved by molecular replacement with PHASER[42] and refined with Refmac[43], SHELXL-2013[44] and COOT[45]. PRODRG was used to build the ligand[46]. Details of the refinement statistics are summarized in Supplementary Table 2.

Data availability. Coordinates and structure factors for the crystal structures of apo and liganded RA95.5-8F were deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) under accession codes PDB 5AOU and PDB 5AN7, respectively.

Received 18 January 2016; accepted 12 July 2016; published online 29 August 2016

References

1. Gefflaut, T., Blonski, C., Perie, J. & Willson, M. Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. *Prog. Biophys. Mol. Biol.* **63,** 301–340 (1995).
2. Dean, S. M., Greenberg, W. A. & Wong, C.-H. Recent advances in aldolase-catalyzed asymmetric synthesis. *Adv. Synth. Catal.* **349,** 1308–1320 (2007).
3. Clapés, P., Fessner, W.-D., Sprenger, G. A. & Samland, A. K. Recent progress in stereoselective synthesis with aldolases. *Curr. Opin. Chem. Biol.* **14,** 154–167 (2010).
4. Windle, C. L., Müller, M., Nelson, A. & Berry, A. Engineering aldolases as biocatalysts. *Curr. Opin. Chem. Biol.* **19,** 25–33 (2014).
5. Mukherjee, S., Yang, J. W., Hoffmann, S. & List, B. Asymmetric enamine catalysis. *Chem. Rev.* **107,** 5471–5569 (2007).
6. Müller, M. M., Windsor, M. A., Pomerantz, W. C., Gellman, S. H. & Hilvert, D. A rationally designed aldolase foldamer. *Angew. Chem. Int. Ed.* **48,** 922–925 (2009).
7. Tanaka, F., Fuller, R. & Barbas III, C. F. Development of small designer aldolase enzymes: catalytic activity, folding, and substrate specificity. *Biochemistry* **44,** 7583–7592 (2005).
8. Wagner, J., Lerner, R. A. & Barbas III, C. F. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science* **270,** 1797–1800 (1995).
9. Barbas III, C. F. *et al.* Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science* **278,** 2085–2092 (1997).
10. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
11. Althoff, E. A. *et al.* Robust design and optimization of retroaldol enzymes. *Protein Sci.* **21,** 717–726 (2012).
12. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9,** 494–498 (2013).
13. Guo, M. T., Rotem, A., Heyman, J. A. & Weitz, D. A. Droplet microfluidics for high-throughput biological assays. *Lab Chip* **12,** 2146–2155 (2012).
14. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16,** 652–656 (1998).
15. Agresti, J. J. *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl Acad. Sci. USA* **107,** 4004–4009 (2010).
16. Baret, J.-C. *et al.* Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip* **9,** 1850–1858 (2009).
17. Kintses, B. *et al.* Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.* **19,** 1001–1009 (2012).
18. Frenz, L., Blank, K., Brouzes, E. & Griffiths, A. D. Reliable microfluidic on-chip incubation of droplets in delay-lines. *Lab Chip* **9,** 1344–1348 (2009).
19. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503,** 418–421 (2013).
20. Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15,** 5631–5640 (1976).
21. Esposito, G. *et al.* Structural and functional analysis of aldolase B mutants related to hereditary fructose intolerance. *FEBS Lett.* **531,** 152–156 (2002).
22. Heine, A. *et al.* Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science* **294,** 369–374 (2001).
23. Henn-Sax, M., Höcker, B., Wilmanns, M. & Sterner, R. Divergent evolution of $(\beta\alpha)_8$-barrel enzymes. *Biol. Chem.* **382,** 1315–1320 (2001).
24. Nagano, N., Orengo, C. A. & Thornton, J. M. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321,** 741–765 (2002).
25. Galkin, A. *et al.* Characterization, kinetics, and crystal structures of fructose-1,6-bisphosphate aldolase from the human parasite, *Giardia lamblia*. *J. Biol. Chem.* **282,** 4859–4867 (2007).
26. Lorentzen, E., Siebers, B., Hensel, R. & Pohl, E. Mechanism of the Schiff base forming fructose-1,6-bisphosphate aldolase: structural analysis of reaction intermediates. *Biochemistry* **44,** 4222–4229 (2005).
27. Lassila, J. K., Baker, D. & Herschlag, D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl Acad. Sci. USA* **107,** 4937–4942 (2010).
28. Meyer, H.-P. *et al.* The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC). *Catal. Sci. Technol.* **3,** 29–40 (2013).
29. Rajagopalan, S. *et al.* Design of activated serine-containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10,** 386–391 (2014).
30. Glasner, M. E., Gerlt, J. A. & Babbitt, P. C. Mechanisms of protein evolution and their application to protein engineering. *Adv. Enzymol. Areas Mol. Biol.* **75,** 193–239 (2007).
31. Garrabou, X., Beck, T. & Hilvert, D. A promiscuous *de novo* retro-aldolase catalyzes asymmetric Michael additions via Schiff base intermediates. *Angew. Chem. Int. Ed.* **54,** 5609–5612 (2015).
32. Garrabou, X., Wicky, B. I. M. & Hilvert, D. Fast Knoevenagel condensations catalyzed by an artificial Schiff-base-forming enzyme. *J. Am. Chem. Soc.* **138,** 6972–6974 (2016).
33. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82,** 447–470 (2013).
34. Taylor, S. V., Kast, P. & Hilvert, D. Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed.* **40,** 3310–3335 (2001).
35. Yang, G. & Withers, S. G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem* **10,** 2704–2715 (2009).
36. Zinchenko, A. *et al.* One in a million: flow cytometric sorting of single cell-lysate assays in monodisperse picolitre double emulsion droplets for directed evolution. *Anal. Chem.* **86,** 2526–2533 (2014).
37. Chen, B. *et al.* High-throughput analysis and protein engineering using microcapillary arrays. *Nat. Chem. Biol.* **12,** 76–81 (2016).
38. Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.* **32,** 1448–1459 (2004).
39. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77,** 61–68 (1989).
40. Zhao, H. & Zha, W. *In vitro* 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nat. Protocols* **1,** 1865–1871 (2006).
41. Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370,** 389–391 (1994).
42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
43. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
44. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
45. Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
46. Schuttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* **60,** 1355–1363 (2004).

Acknowledgements

We thank C. Stutz-Ducommun and B. Blattmann from the Protein Crystallization Core Facility at the University of Zurich and the staff at the Swiss Light Source (Paul Scherrer Institute) for outstanding technical support, as well as E. Mayot for surfactant synthesis. We are grateful to the mass spectrometry service of the Laboratory of Organic Chemistry at ETHZ and the Protein Analysis Group at the Functional Genomics Center of the University of Zurich. We thank A. Green, R. Zschoche, M. Reichen, R. Calbrix and all the members of the Hilvert and Griffiths labs for fruitful discussions. This work was generously supported by the Swiss National Science Foundation (D.H.).

Author contributions

D.H., A.D.G., D.B., P.R.E.M., X.G., A.G. and R.O. designed the experiments. R.O. and A.G. developed the assay and evolved retro-aldolases. R.O. and X.G. biochemically characterized the variants. X.G. performed biocatalytic aldol reactions and the synthesis of standards. P.R.E.M. and R.O. crystallized the proteins and solved their structures. The manuscript and figures were prepared by R.O., X.G., A.D.G. and D.H.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.D.G. and D.H.

Competing financial interests

The authors declare no competing financial interests.