nature chemistry

ARTICLES

PUBLISHED ONLINE: 22 JANUARY 2018 | DOI: 10.1038/NCHEM.2927

Evolving artificial metalloenzymes via random mutagenesis

Hao Yang[1†], Alan M. Swartz[2†], Hyun June Park[2†], Poonam Srivastava[3†], Ken Ellis-Guardiola[2], David M. Upp[2], Gihoon Lee[2,4], Ketaki Belsare[2], Yifan Gu[2], Chen Zhang[5], Raymond E. Moellering[2,4] and Jared C. Lewis[2*]

Random mutagenesis has the potential to optimize the efficiency and selectivity of protein catalysts without requiring detailed knowledge of protein structure; however, introducing synthetic metal cofactors complicates the expression and screening of enzyme libraries, and activity arising from free cofactor must be eliminated. Here we report an efficient platform to create and screen libraries of artificial metalloenzymes (ArMs) via random mutagenesis, which we use to evolve highly selective dirhodium cyclopropanases. Error-prone PCR and combinatorial codon mutagenesis enabled multiplexed analysis of random mutations, including at sites distal to the putative ArM active site that are difficult to identify using targeted mutagenesis approaches. Variants that exhibited significantly improved selectivity for each of the cyclopropane product enantiomers were identified, and higher activity than previously reported ArM cyclopropanases obtained via targeted mutagenesis was also observed. This improved selectivity carried over to other dirhodium-catalysed transformations, including N–H, S–H and Si–H insertion, demonstrating that ArMs evolved for one reaction can serve as starting points to evolve catalysts for others.

Selective catalysis can greatly improve the efficiency of chemical synthesis, but designing and optimizing catalysts with high activity and selectivity rank among the most challenging endeavours in synthetic chemistry[1]. The catalytic proficiency of enzymes acting on their native substrates has inspired a range of approaches to engineer enzymes for chemical synthesis[2]. The introduction of mutations to improve enzyme activity and selectivity toward substrates and reactions of interest has proven particularly powerful, but these efforts are complicated by our inability to fully predict how a given mutation will alter enzyme function[3]. In the absence of this understanding and given the astronomical number of mutations that could be introduced into an enzyme, efficient means of introducing and identifying mutations that provide improved function are required.

When available, information regarding enzyme mechanism, structure or sequence homology has proven useful in helping to identify beneficial mutations in an enzyme, typically introduced within its active site[4,5]. Such information may not be available, however, or it may prove insufficient to provide the level of improvement needed for a desired application. More fundamentally, targeted mutagenesis efforts aimed at the active site of an enzyme miss the potential benefits of mutations that are not apparent from available data or that emerge from synergistic mutations throughout the protein scaffold[3]. To confront this challenge, Arnold and Stemmer independently developed approaches to engineer enzymes by incorporating random mutations, screening or selecting for improved enzyme variants, and iterating the process until the desired level of improvement is obtained[6,7]. Directed evolution, as this process is called, requires no detailed knowledge of enzyme structure or mechanism and can be used to identify beneficial mutations throughout an enzyme sequence[3]. While powerful, directed evolution of enzymes with useful properties typically requires many iterations of mutagenesis and screening. Hybrid directed evolution approaches have therefore been developed to combine the benefits of targeting mutations (or sets of mutations) based on available information and the benefits of sampling diverse mutations throughout the enzyme structure[2,8]. Regardless of the methods used, introducing diversity throughout enzyme structures has proven critical for engineering efficient enzyme activity toward both native and non-native substrates.

The success of protein engineering for biocatalysis has inspired similar efforts to engineer artificial metalloenzymes (ArMs), hybrid catalysts composed of protein scaffolds containing synthetic metal cofactors[9]. Unlike natural enzymes, however, extensive mechanistic data on these systems are not available (they must be created to begin with), and it is not obvious that the phylogeny of a scaffold protein, which reflects its natural evolution for a native function, will provide insight into mutations that could improve ArM activity[8,10]. Although structural information can be used, it is not always available, and homology models are imperfect structural references. Even if the structure of a given scaffold is known, solving structures of descendent ArMs may not be possible or informative, because the flexibility of many synthetic cofactor linkers (for example, **1** (Fig. 1c)) can lead to poor occupancy in electron density maps. Despite these limitations, crystal structures of parental scaffolds have guided all ArM engineering efforts reported so far. For example, highly selective ArMs have been obtained via structure-guided, targeted mutagenesis of scaffold proteins, including streptavidin[11,12], cyt cb_{562} (ref. 13) and myoglobin[14–17].

We previously applied a targeted mutagenesis approach to engineer dirhodium ArMs that catalyse enantioselective cyclopropanation of styrenes (Fig. 1a)[18]. This effort involved replacing the active site

---

[1]Merck Research Laboratories, 126 E. Lincoln Avenue, Rahway, New Jersey 07065, USA. [2]Department of Chemistry, University of Chicago, 5735 S. Ellis Avenue, Chicago, Illinois 60637, USA. [3]Cargill, Inc., 14800 28th Avenue N, Plymouth, Minnesota 55447, USA. [4]Institute for Genomics and Systems Biology, University of Chicago, 929 E. 57th Street, Chicago, Illinois 60637, USA. [5]Provivi, Inc., 1701 Colorado Avenue, Santa Monica, California 90404, USA. †These authors contributed equally to this work. *e-mail: jaredlewis@uchicago.edu

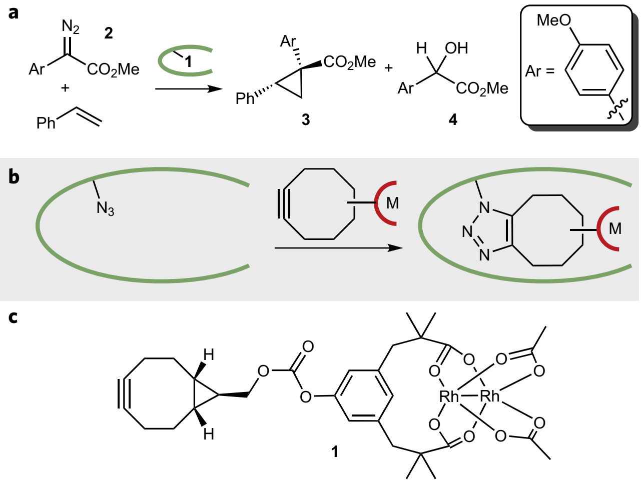


Figure 1 | Model reaction and ArM structure. **a**, Model cyclopropanation reaction used for ArM evolution. The dirhodium cofactor catalyses diazo decomposition to generate a rhodium carbenoid intermediate that can insert into the olefin π bond or (because the reactions are carried out in aqueous solution) the water O–H bond. The protein scaffold provides chemoselectivity and enantioselectivity. Reactions were conducted using 22 mM styrene, 4.4 mM diazo and 1 mol% ArM in 10% vol/vol THF/50 mM PIPES (pH 7.4) containing 1.75 M NaBr at 4 °C. **b**, ArM formation via SPAAC[19] between bicyclononyne-substituted metal complexes and a genetically encoded 4-L-azidophenylalnine residue, allowing covalent attachment of the cofactor, even in cellular lysate. **c**, Structure of dirhodium cofactor **1**.

serine (S477) of *Pyrococcus furiosus* (*Pfu*) prolyl oligopeptidase (POP) with a genetically encoded *p*-azidophenylalanine (Z) residue to which cofactor **1** was linked via strain-promoted azide–alkyne cycloaddition (SPAAC, Fig. 1b,c). In contrast to the scaffolds noted above, an X-ray crystal structure for *Pfu* POP had not been reported at the outset of our studies, it does not bind metal cofactors that can be substituted to form ArMs, and its native substrate binding was not used for ArM formation. Instead, *Pfu* POP was selected on the basis of its high stability[19] and the large active site common to POP family enzymes[20], which earlier efforts[21] suggested would be necessary to contain and impart selectivity to **1**. A previously reported homology model of this enzyme provided some guidance for targeting mutations[22], but a large number of individual mutants had to be cloned and evaluated[18], which proved to be a time-consuming effort due to differences between the actual and homology structures and a complete lack of information regarding cofactor location within the active site.

The tedious sequential mutagenesis required to generate selective POP ArMs reflected limitations on the extent to which information-based targeted mutagenesis approaches commonly used for engineering natural enzymes[10] would be effective in ArM engineering. We reasoned that random mutagenesis would be a powerful tool for ArM engineering because it does not require structural information and could reveal beneficial mutations distal to ArM active sites[3]. Although such mutations have proven highly important for engineering natural enzymes, they are difficult to predict and their potential to impact ArM catalysis has not been addressed[15]. From the perspective of transition-metal catalysis, random mutagenesis of ArMs also offers the unique opportunity to probe the effects of structural changes far beyond the metal secondary coordination sphere. ArM evolution via random mutagenesis is complicated, however, by challenges associated with high-throughput expression of scaffold libraries, introducing synthetic metal cofactors into these libraries, eliminating background catalysis by free cofactor, and ensuring that cellular lysate components do not interfere with catalysis[12,23,24]. To address these challenges, here we used a dirhodium ArM cyclopropanase previously engineered in our laboratory as a platform to explore the potential to improve ArM selectivity via random mutagenesis. Libraries generated via both random mutagenesis of the entire POP β-propeller domain and combinatorial mutagenesis[22] of residues projecting into the POP active site were examined. High selectivity for either product enantiomer was achieved from mutations both inside and outside the ArM active site, and the evolved cyclopropanases had improved selectivity toward a number of additional dirhodium-catalysed reactions, including N–H insertion, S–H insertion and Si–H insertion. These results highlight the potential for random mutagenesis to identify mutations in ArMs that would be difficult to predict even

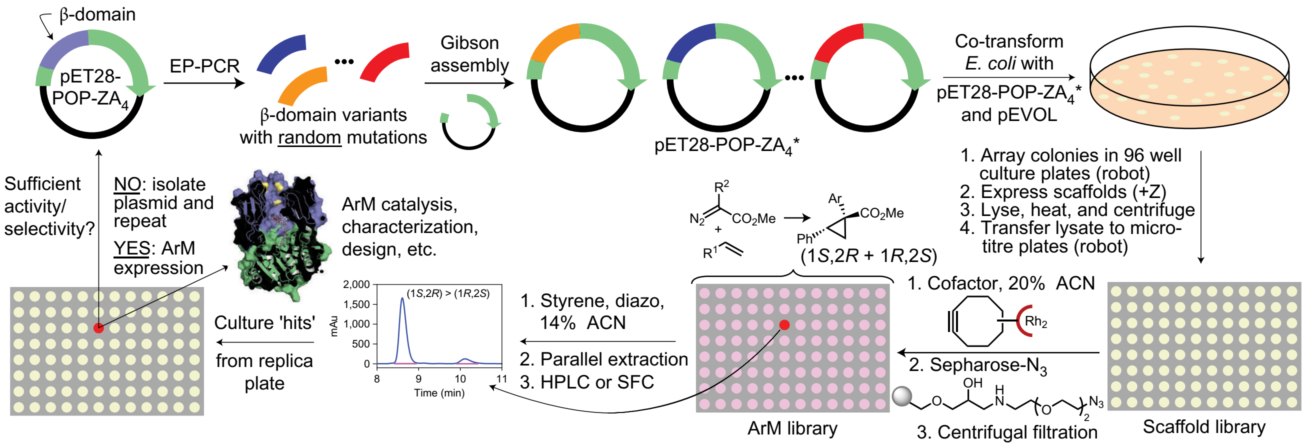


Figure 2 | Overview of ArM evolution protocol. From top left: The gene encoding a POP scaffold (for example, 0-ZA$_4$) in pET-28 (the POP β-domain is shown in purple in both plasmid and the ArM structure model) is used to generate a library of β-domain variants with random mutations via error-prone (EP) PCR. The remaining backbone is amplified separately, and the gene library is ligated via Gibson assembly. The gene library is co-transformed into *E. coli* with pEVOL-pAzF, a plasmid containing an orthogonal tRNA and aaRS for Amber stop codon suppression. A colony picker robot is used to array colonies into 96-well plates, where the POP genes are expressed and cells are lysed, heated and centrifuged. Next, a liquid handling robot is used to transfer the lysate to a fresh 96-well plate, cofactor is added to generate POP ArMs, and an azide-substituted resin is added to scavenge unreacted cofactor. The ArM library is screened by HPLC or supercritical fluid chromatography (SFC) for increased enantioselectivity of olefin cyclopropanation. Putative hits are isolated and cultured for verification on a larger scale without the presence of cellular lysate debris. If the putative hit is validated, it is used as parent for an additional round of mutagenesis until the desired activity/selectivity is observed.

with detailed structural data. When combined with existing targeted mutagenesis approaches[8,10], our strategy provides a general framework for ArM evolution analogous to that used for natural enzymes[2], and suggests that similar improvements in the efficiency of diverse ArMs should be possible.

Results

ArM evolution via error-prone PCR. Our ArM evolution efforts relied on several unique aspects of the *Pfu* POP scaffold[18,22] and the SPAAC bioconjugation method used for cofactor incorporation (Fig. 2)[21]. At the outset of our efforts, a crystal structure of *Pfu* POP was not available, so error-prone PCR was used to introduce mutations throughout the POP β-propeller domain (Q48 to V335, Fig. 2) that comprises the putative ArM active site[18]. The high stability of POP allowed the use of high mutation rates (typically four to five residue mutations per variant)[25] and various manipulations involved in parallel ArM formation. Gibson assembly of the β-propeller variants provided the desired scaffold libraries without the need for introducing restriction sites in the β-propeller domain. Conditions were then optimized for expressing POP in *Escherichia coli* in high yield in 24- or 96-well plates. The stability of POP also allowed for thermal denaturation of *E. coli* proteins following cell lysis[26]. Centrifugation provided cell lysates containing $\sim$50 μM POP based on SDS–PAGE analysis (Supplementary Fig. 1). The high efficiency of SPAAC allowed for rapid bioconjugation of **1** to POP in 96-well plates using a roughly twofold excess of cofactor (93.75 μM) over the average amount of protein in each well (Supplementary Fig. 1). A commercially available azide-substituted resin could be used to scavenge the excess cofactor following bioconjugation, and a comparable resin was prepared to facilitate this process[27]. In the absence of this step, non-selective reactions catalysed by free cofactor will compete with ArM catalysis, reducing the observed selectivity. Cyclopropanation reactions were then conducted in deep-well plates, the reaction mixtures were extracted with hexanes, and the organic extracts were analysed by HPLC.

This procedure was used to evolve the selectivity of POP variant 0-ZA$_4$, the starting point for our rational design effort (previously called POP-ZA$_4$-**1**)[18] for the cyclopropanation of 4-methoxystyrene (Fig. 1a). Gratifyingly, only three rounds of mutagenesis and screening (96, 48 and 576 variants per round, respectively; Supplementary Fig. 2) were required to obtain 92% e.e. for the target cyclopropanation reaction using ArM variant 3-VRVH (Fig. 3a and Supplementary Table 2). Variants with improved selectivity relative to the parent enzyme assayed in the same plate were taken to be hits, and these putative hits were validated following purification. The variant with the highest selectivity was selected as the parent for the next round of mutagenesis (Supplementary Fig. 3). The high mutation rates used effectively allowed for multiplexed analysis of several mutations in each variant, and the most improved variant in each round contained four residue mutations (Fig. 3). Analysis of product formation over time in reactions catalysed by 3-VRVH, 0-ZA$_4$ and HFF, the latter of which was generated during our previous rational design effort[18], revealed that 3-VRVH possessed significantly higher activity than the unevolved or rationally designed ArMs (Fig. 3b and Supplementary Fig. 6).

Individually reverting each mutation in each of the improved variants indicated that only three mutations (S301G/G99S/Y326H) were required for most of the improvement in selectivity observed in 3-VRVH, which contains 12 mutations relative to 0-ZA$_4$ (Fig. 3 and Supplementary Fig. 7). On the other hand, the product yield obtained for the variant containing only these essential mutations, 1-GSH, was significantly reduced (Fig. 3, inset). This finding suggests that increased selectivity appears to have evolved at the expense of activity, but other mutations throughout the POP structure overcome this limitation, leading to higher conversion for 3-VRVH relative to 1-GSH.

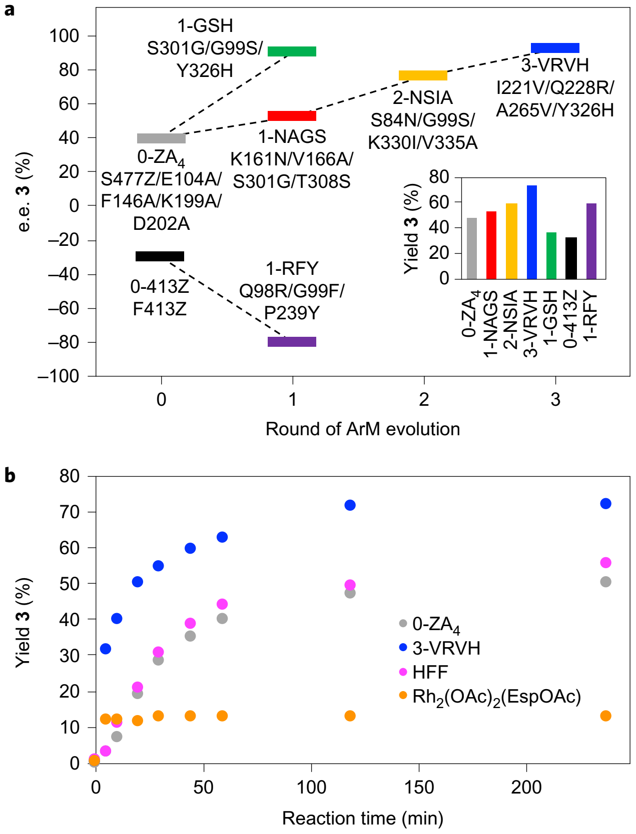


Figure 3 | Overview of the directed evolution lineages generated and time-course comparison of several catalysts. a, Enantioselectivities and yields (inset) for cyclopropanation of styrene catalysed by evolved ArM variants to give **3** (Fig. 1a). Although e.e. was the screening criterion, yields of the desired product also increased across lineages. Each variant contains the mutations indicated, plus those from the previous round(s) of evolution. The residues identified from deconvolution experiments were cloned into a minimal mutant (1-GSH), which was able to provide equivalent enantioselectivity to the final mutant (3-VRVH), but significantly lower yield. This highlights the complexity of biomolecule scaffolds and the importance of random mutagenesis for ArM engineering. Reactions were conducted as shown in Fig. 1a using 22 mM styrene, 4.4 mM diazo and 1 mol% catalyst in 10% vol/vol THF/50 mM PIPES (pH 7.4) containing 1.75 M NaBr or NaCl at 4 °C for 4 h. e.e. and yield of **3** were determined by analysis of HPLC chromatograms for crude reaction mixtures relative to internal standards. **b**, Time-course experiments for reactions catalysed by different dirhodium catalysts. The yield and qualitative rate for the evolved variant 3-VRVH exceeds parent 0-ZA$_4$ as well as a previously reported ArM produced via rational design. The reaction in aqueous buffer of the free cofactor is also shown. All data points shown are an average of two reactions.

During the course of our evolution efforts, we solved the crystal structure of wild-type (wt) *Pfu* POP (PDB ID 5T88), which allowed for structural analysis of the mutations identified via directed evolution. Of the mutations in 1-GSH, G99S and Y326H are in the POP active site while S301G is not (Fig. 4). Notably, G99F and L328H were identified as beneficial mutations in our previous engineering effort, but the crystal structure showed that the locations of these residues differed significantly from those suggested by the homology model used[18]. The identification of distinct mutations via targeted and random mutagenesis highlights the utility of both approaches for ArM engineering and the range of different active site configurations that can improve ArM selectivity.

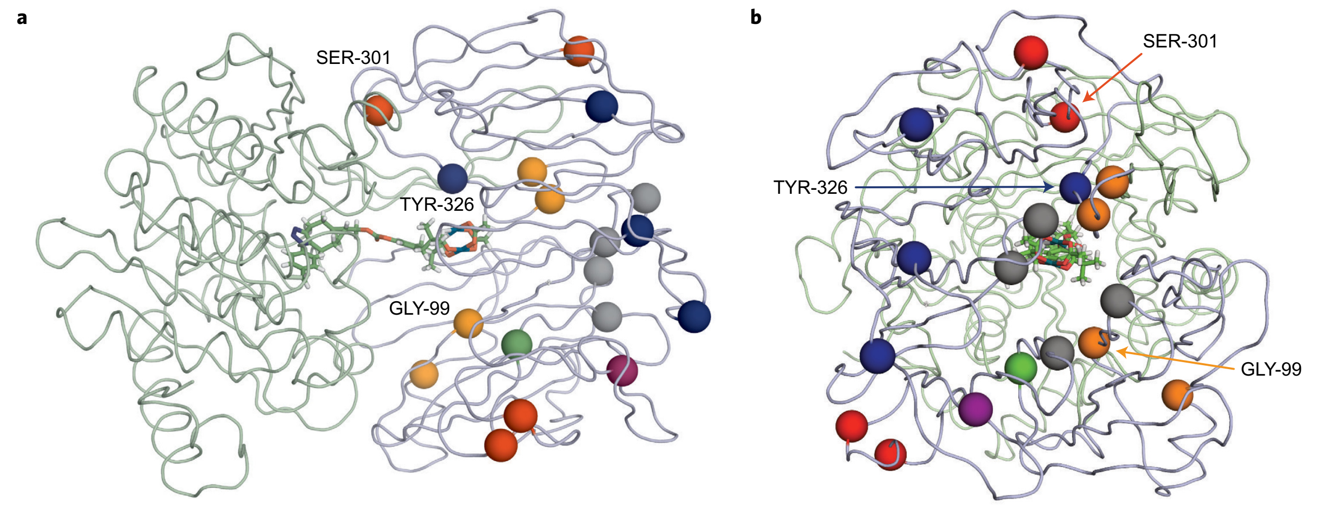


Figure 4 | Location of mutations in evolved ArMs. a,**b**, Side (**a**) and top (**b**) views of the POP ribbon model, with spheres showing the location of mutations in variants 0-ZA$_4$ (grey), 1-NAGS (red), 2-NSIA (orange) and 3-VRVH (blue). Mutations with greatest impact on enantioselectivity are labelled. Residues modified via carbene insertion (*vide infra*) are also shown as spheres (purple, W175; green, W142).

Moreover, the improvement in selectivity resulting from S301G and the improvements in yield resulting from mutations throughout the POP scaffold clearly attest to the importance of non-active site mutations for ArM optimization and library methods that enable their identification.

ArM evolution via combinatorial codon mutagenesis. Despite the efficiency with which directed evolution was able to improve cyclopropanation enantioselectivity, at no point was a variant that provided significant selectivity for the opposite product enantiomer observed. Previous reports have shown that altering the site of cofactor linkage within a scaffold can alter ArM selectivity[28], so we used the *Pfu* POP crystal structure to identify alternative linkage sites for **1**. POP variant F413Z (Fig. 5a, blue sphere) led to an ArM that provided modest selectivity for the desired product enantiomer (30% e.e., Fig. 3), opposite that obtained using the S477Z linkage site (Fig. 5a, green sphere). While the *Pfu* POP crystal structure does not reveal specific residues that could be targeted to further improve the selectivity of variant F413Z (ref. [8]), it does indicate which residues actually project into the POP active site. We therefore reasoned that random mutagenesis of these active site residues might be a useful approach for rapidly improving the selectivity of variant F413Z.

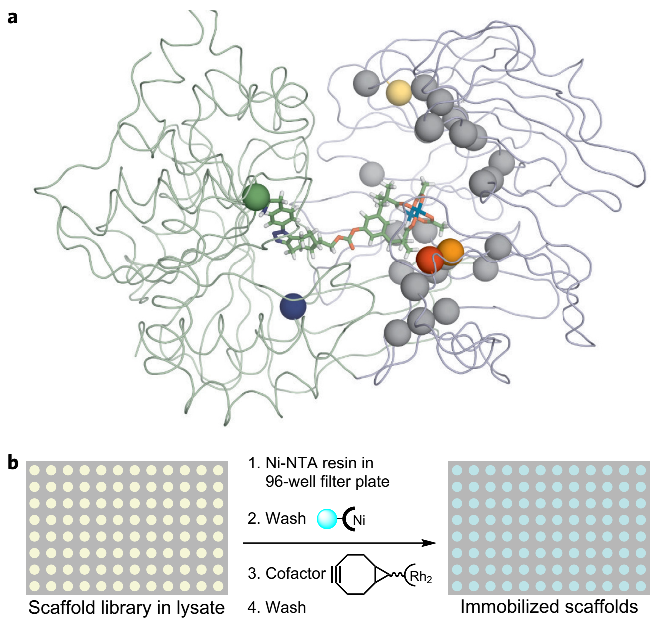


Figure 5 | Combinatorial codon mutagenesis sites and protocol. a, Ribbon model of POP variant 0-ZA$_4$ showing the location of residue 477 (green sphere), residue 413 (blue sphere) and residues selected for combinatorial codon mutagenesis (remaining spheres). Mutations in 1-RFY (Q98R, G99F and P239Y) are shown as red, orange and yellow spheres, respectively. **b**, Scaffold immobilization and ArM formation on Ni-NTA resin in 96-well plates using libraries of scaffold variants generated via error-prone PCR or combinatorial codon mutagenesis (CCM) as shown in Fig. 2.

Based on this hypothesis, a combinatorial codon mutagenesis library[22] of POP-413Z variants, each containing an average of two degenerate NDT codons at any of the 25 remaining active site residues, was constructed (Fig. 5a). The use of the NDT codon eliminated the possibility of stop codon mutations and allowed for rapid and efficient sampling of random mutations throughout the POP active site. Ninety-two members of this library were screened on-bead using a modified version of the protocol shown in Fig. 2 in which POP-413Z variants in cell lysate were immobilized on Ni-NTA (nitrilotriacetic acid) resin[29] preloaded into 96-well filter plates (Fig. 5b and Supplementary Fig. 8). ArM formation and catalysis using immobilized POP-413Z simplified cofactor removal, reaction set-up and product isolation relative to the analogous processes in solution. Control reactions using lysate containing POP-413A or lacking POP entirely provided a conversion of 2–6% and negligible enantioselectivity, indicating that properly assembled ArM was responsible for catalysis. One variant identified via this process, 1-RFY (POP-413Z Q98R/G99F/P239Y), possessed significantly improved enantioselectivity, providing the desired cyclopropanation enantiomer in 80% e.e. (Fig. 3 and Supplementary Fig. 9; see Supplementary Fig. 10 for hit validation under screening conditions). Individually reverting each mutation in 1-RFY led to decreased product yield and enantioselectivity (Supplementary Table 6), indicating that all three mutations contributed to the improved selectivity of 1-RFY. Notably, neither Q98 nor P239 was targeted in our earlier more conservative targeted mutagenesis efforts, and no beneficial mutations were identified at these sites during the directed evolution campaign targeting the original enantiomer. These data show how combinatorial codon mutagenesis uniquely balances the benefits of sampling random mutations at a large number of sites with those of restricting those sites to particular regions of an enzyme[10]. Although more focused mutagenesis strategies can often be used for engineering natural enzymes[8], these require sequence and structural information that is often not available for ArMs. Our combined approach used here is an attractive strategy for these enzymes.

***In situ* ArM modification via carbene insertion.** To begin to shed light on the mechanisms by which the POP scaffolds evolved in this work impart selectivity to synthetic cofactor **1**, reaction profiles for the ArMs in the 3-VRVH lineage were examined. These profiles showed significant differences in both cyclopropanation yield and enantioselectivity depending on the reaction times used (Fig. 6). Interestingly, the e.e. of the cyclopropane formed was observed to decrease during the course of these reactions, and this decrease itself appeared to decrease for variants along the 3-VRVH lineage. Although the free energy change associated with the observed decrease in enantioselectivity of 3-VRVH (from 95% e.e. to 92% e.e.) is roughly equivalent to that associated with the decrease in enantioselectivity of ZA_4 (from 65% e.e. to 50% e.e.), 1-GSH undergoes no change in enantioselectivity. Moreover, analysis of reactions catalysed by 0-ZA_4 and 3-VRVH starting immediately after ArM addition (Supplementary Fig. 6) revealed only a negligible decrease in e.e. for 3-VRVH (from 94% e.e. to 93% e.e.) and a larger decrease in e.e. for 0-ZA_4 (from 73% e.e. to 40% e.e.). The former shows that minor variation in 3-VRVH selectivity is significantly impacted by error associated with the HPLC measurements used to determine e.e., while the latter presumably reflects rapid loss of enantioselectivity at times before those initially analysed in Fig. 6. Based on these observations and a number of studies on dirhodium-catalysed carbene insertion into peptides[30], we hypothesized that modification of the POP scaffold via carbene insertion could be occurring in certain POP variants. If this were the case, it would provide some rationale by which residues could be targeted to improve ArM efficiency. Indeed, Arnold recently reported that carbene insertion into cytochrome P450 BM3 during cyclopropanation reactions catalysed by this enzyme led to decreased cyclopropanation activity and selectivity[31].

To determine if carbene insertion into POP residues was leading to the observed decrease in POP ArM cyclopropanase selectivity (Fig. 6), both 0-ZA_4 and 3-VRVH were purified following incubation with styrene and diazo **2** (Fig. 1a), double-digested using cyanogen bromide, followed by trypsin, and analysed by tandem LC–MS/MS (Supplementary Fig. 12). The double-digestion procedure enabled near complete sequence coverage of the thermostable protein (trypsin alone was insufficient), which was exploited for unbiased differential modification searches for carbene-modified residues. These studies identified two modification sites, W175 and W142 (Fig. 4: purple sphere, W175; green sphere, W142), that were exclusively detected under carbene-generating conditions and not control reactions of scaffold alone (Supplementary Fig. 12). Ion intensities of both the modified and unmodified tryptic peptides indicated a low level of modified W175, which is a surface residue (Supplementary Fig. 12). In contrast, extensive carbene insertion was detected at W142 in both 1-NAGS (from the first round of ArM evolution) and 3-VRVH (Supplementary Fig. 12). Both of these residues in variant 0-ZA_4 were individually mutated to alanine, but both 0-ZA_4-W175A and -W142A provided yield and selectivity profiles similar to those of 0-ZA_4 (that is, similar decreases in e.e. over time, Supplementary Fig. 12). These data indicate that, while scaffold modification occurs during catalysis, modification at W142 or W175 is not responsible for the decrease in POP ArM cyclopropanase selectivity observed.

Comparable levels of modification for evolved variants in the cyclopropanase lineage were also observed based on the relative ion counts of modified and unmodified fragments and intact scaffold observed via LC–MS/MS (Supplementary Fig. 12). This finding does not, however, rule out the possibility that low levels of carbene modification at residues not detected in our analysis could impact ArM selectivity. Given the wide range of residues that can be modified via proximity-driven carbene insertion reactions[32], removing all such residues from an ArM active site will often be challenging and would greatly constrain the range of residues that could be used to improve ArM selectivity. The fact that 3-VRVH suffers from only a minor loss in enantioselectivity over time suggests that if carbene modifications at sites not detected in our analysis impact ArM selectivity, random mutations identified via directed evolution reduced this impact. Indeed, the ability to evolve ArMs via iterative random mutagenesis provides a general framework for improving ArM selectivity, even when the molecular origins of selectivity are unknown, just as can be accomplished for natural enzymes[3].

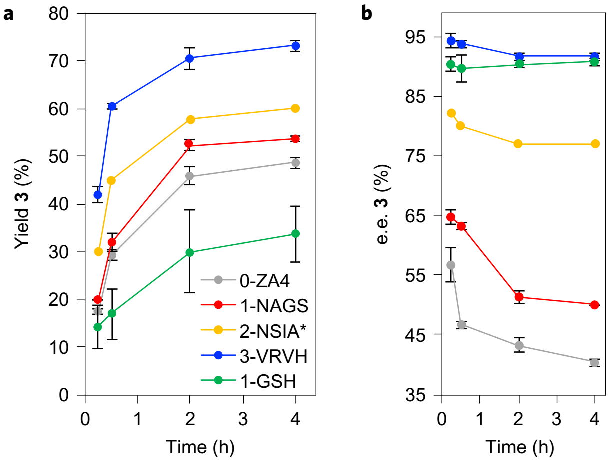
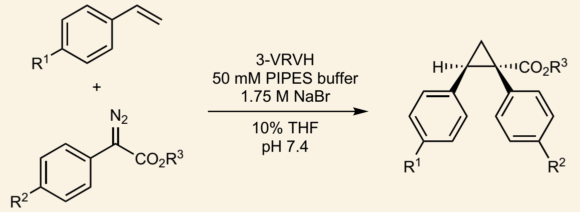


Figure 6 | Time-course experiments of ArM-catalysed cyclopropanations of styrene with (4-methoxyphenyl)methyldiazoacetate. See Supplementary Table 2 for full data. **a**, Plot of cyclopropane yield versus time, showing increases for each generation of mutants. **b**, Plot of cyclopropane e.e. versus time. A decrease in percent e.e. over the course of the reaction was observed, which decreases along the lineage. The cause of this was investigated but not identified (see main text). Reactions were performed in triplicate with standard deviation shown. *2-NSIA reactions were performed in duplicate, prohibiting standard deviation calculation.

Substrate and reaction scope of evolved ArMs. ArM evolution also provides lineages of improved variants that can be studied to provide insight into ArM mechanism and function. Given the wide range of reactions catalysed by dirhodium complexes, one immediate question that might be asked of the ArM cyclopropanase lineage described in this work is the extent to which improved selectivity for styrene cyclopropanation carries over to other dirhodium-catalysed reactions. We previously showed that our rationally designed ArM variants provided improved enantioselectivity on several additional styrene/diazo pairs, and this proved to be the case for the variants evolved in the current study (Table 1). Moreover, improved enantioselectivity was also observed for other dirhodium catalysed reactions, including formal carbene insertion into Si–H, S–H and N–H bonds (Table 2). 1-RFY provided the opposite enantiomer of each of these products, although lower selectivity was observed (Supplementary Table 3). These results illustrate the power of directed evolution to provide ArMs for reactions beyond the scope of the initial evolution target (that is, enantioselective cyclopropanation)[33]. Nonetheless, the general improvement observed for these reactions, which proceed via different mechanisms in solution, suggests that even the evolved 3-VRVH and 1-RFY scaffolds represent only the tip of the iceberg

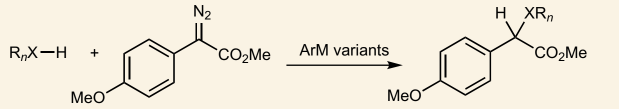
Table 1 | Additional products from 3-VRVH catalysed (1 mol%) cyclopropanation using conditions from Fig. 1a.

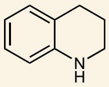


Entry	R^1	R^2	R^3	Yield (%)	e.e. (%)
1	MeO	MeO	Me	25	86
2	H	MeO	Et	43	94
3	Cl	MeO	Me	46	86
4	H	H	Me	66	88
5	H	Cl	Me	52	90
6	H	Br	Me	54	92

Reactions conducted as shown in Fig. 1a using 22 mM styrene, 4.4 mM diazo and 1 mol% ArM in 10% vol/vol THF/50 mM PIPES (pH 7.4) containing 1.75 M NaBr at 4 °C for 4 h. Yields and e.e. were determined by HPLC using an internal standard (1,3,5-trimethoxybenzene).

Table 2 | Additional X-H insertion reactions catalysed by different POP ArM variants (1 mol%).



Entry	Substrate	Yield, % (e.e.)		
		0-ZA$_4$	1-NAGS	3-VRVH
1	PhMe$_2$Si–H	45 (6)	43 (14)	35 (64)
2	PhS–H	51 (6)	66 (10)	64 (32)
3		55 (8)	77 (6)	73 (40)

Reactions conducted as shown in Fig. 1a using 22 mM styrene, 4.4 mM diazo and 1 mol% ArM in 10% vol/vol THF/50 mM PIPES (pH 7.4) containing 1.75 M NaBr at 4 °C for 4 h. Yields and e.e. were determined by HPLC using an internal standard (1,3,5-trimethoxybenzene).

of reactivity that may be accessed using dirhodium ArMs. Further evolution aimed at improving both the selectivity and reaction specificity of these ArMs will shed light on the extent to which proteins can control the reactivity of synthetic organometallic complexes and the mechanism(s) by which this control is imparted.

Summary and conclusion. Random mutagenesis has proven to be a critical tool for engineering natural enzymes for selective catalysis[3]. In the absence of a comprehensive understanding of enzyme structure–activity relationships, the ability to create and survey the impact of diversity throughout an enzyme's structure is essential to identifying beneficial mutations that can, collectively, lead to enormous improvements in catalytic efficiency. Many reports have shown that ArM function can be significantly improved via mutations targeted to ArM active sites[15]. Before this report, however, random mutagenesis had not been used for ArM evolution, and the question of whether ArM function could be further (or even significantly) improved via mutations at sites throughout the ArM structure had not been addressed.

We have shown that iterative random mutagenesis can be used to generate dirhodium ArM variants with improved enantioselectivity for a model styrene cyclopropanation reaction. The evolved ArMs also exhibited improved selectivity toward other dirhodium-catalysed reactions including Si–H insertion and formal insertion into N–H and S–H bonds. These improvements resulted primarily from higher inherent ArM selectivity, as evidenced by increased enantioselectivity at early reaction times, and may also reflect reduced degradation of enantioselectivity during the course of catalysis (Fig. 4). Importantly, mutations throughout the POP scaffold, including at sites distal to the active site, significantly impacted both ArM activity and selectivity. This finding has broad implications for metalloprotein design. The vast majority of efforts reported so far focus on engineering metal binding sites[34]. Using the approach described herein, the primary metal binding site was defined by the cofactor used, and only secondary sphere changes were made. Moreover, most of these changes were distal to the pocket in which the metal centre resides, clearly showing that a focus on the entire protein, rather than just the primary coordination sphere or even residues proximal to it, can lead to significant improvements over current approaches.

As is frequently found for mutations identified via directed evolution of natural enzymes, the mechanisms by which these distal mutations impact ArM selectivity are not yet clear. LC–MS/MS analysis of ArMs in the selectivity lineage following cyclopropanation reactions revealed significant scaffold modification via carbene insertion, but ArM selectivity still decreased when modified residues were replaced with alanine, ruling out scaffold modification as a major factor for improved ArM selectivity. Distal mutations have often been suggested to affect conformational changes required for enzyme catalysis[35] and, indeed, conformational dynamics are proposed to play a significant role in the native peptide hydrolysis reactions catalysed by POP[36]. The impact of mutations identified in this work on both POP hydrolase and ArM cyclopropanase activity are thus the subject of ongoing efforts in our group. We believe such studies will provide valuable insights into how proteins can control the reactivity of synthetic metal centres, and the POP platform discussed in this work will provide a valuable subject for these studies. Nonetheless, the improved catalytic efficiency of ArM variants resulting from mutations throughout the POP scaffold, including distal to the active site, clearly demonstrates the utility of random mutagenesis for ArM optimization. Although we used random mutagenesis to identify these mutations, any number of mutagenesis schemes could ultimately be used to achieve similar results[2]. We anticipate that similar efforts with other ArMs, particularly those generated from scaffold proteins that did not evolve in nature to bind metal cofactors[14], will enable the same levels of improvement for these synthetic systems as has been observed for the evolution of natural enzymes[3].

Methods

POP scaffold expression and bioconjugation to generate the corresponding ArMs were conducted as previously described[13]. Solutions of aryldiazoacetate (25 μl, 96 mM, in THF), styrene (25 μl, 485 mM, in THF) and ArM (500 μl, 48 μM, in 50 mM PIPES buffer, pH 7.4, with 1.75 M NaBr or NaCl additive) were added to a 1.5 ml microcentrifuge tube. The final concentrations of the reagents were 22 mM olefin, 4.4 mM aryldiazoacetate and 44 μM ArM. The resulting mixture was left shaking at 750 r.p.m. on an Eppendorf Thermomixer R at 4 °C for 4 h. The reaction was quenched by adding 20 μl 1,4-benzodioxole solution (22 mM, in THF) and 600 μl ethyl acetate. The mixture was vortexed and centrifuged (15,000*g*, 3 min). The top organic layer was collected and the bottom aqueous layer was extracted with 600 μl ethyl acetate, twice. The organic layers were combined, evaporated and redissolved in 200 μl 10% vol/vol 2-propanol in hexane. This solution (5 μl) of the crude product was analysed on normal-phase HPLC to determine product yield and e.e. Detailed procedures for the directed evolution method used to improve e.e. and ArM characterization are provided in the Supplementary Information.

Data availability. Complete experimental procedures, including primer sequences, synthetic procedures, characterization data, library generation and screening protocols are described in the Supplementary Information. Sequencing and library screening data are available upon request. The crystal structure for POP R464L has been deposited in the Protein Data Bank (PDB) under accession no. 5T88.

Received 14 March 2017; accepted 30 November 2017; published online 22 January 2018

References

1. Mahatthananchai, J., Dumas, A. M. & Bode, J. W. Catalytic selective synthesis. *Angew. Chem. Int. Ed.* **51,** 10954–10990 (2012).
2. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
3. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10,** 866–876 (2009).
4. Lu, Y., Berry, S. & Pfister, T. Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. *Chem. Rev.* **101,** 3047–3080 (2001).
5. Gerlt, J. A. & Babbitt, P. C. Enzyme (re)design: lessons from natural evolution and computation. *Curr. Opin. Chem. Biol.* **13,** 10–18 (2009).
6. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl Acad. Sci. USA* **90,** 5618–5622 (1993).
7. Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370,** 389–391 (1994).
8. Lutz, S. Beyond directed evolution–semi-rational protein engineering and design. *Curr. Opin. Biotech.* **21,** 734–743 (2010).
9. Lewis, J. C. Artificial metalloenzymes and metallopeptide catalysts for organic synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
10. Kazlauskas, R. J. & Bornscheuer, U. T. Finding better protein engineering strategies. *Nat. Chem. Biol.* **5,** 526–529 (2009).
11. Reetz, M. T., Peyralans, J. J. P., Maichele, A., Fu, Y. & Maywald, M. Directed evolution of hybrid enzymes: evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chem. Commun.* **2016,** 4318–4320 (2006).
12. Jeschek, M. *et al.* Directed evolution of artificial metalloenzymes for *in vivo* metathesis. *Nature* **537,** 661–665 (2016).
13. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **346,** 1525–1528 (2014).
14. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534,** 534–537 (2016).
15. Hyster, T. K. Genetic optimization of metalloenzymes: enhancing enzymes for non-natural reactions. *Angew. Chem. Int. Ed.* **55,** 7344–7357 (2016).
16. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal substitution modulates the reactivity and extends the reaction scope of myoglobin carbene transfer catalysts. *Adv. Synth. Catal.* **359,** 2076–2089 (2017).
17. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Stereoselective olefin cyclopropanation under aerobic conditions with an artificial enzyme incorporating an iron-chlorin e6 cofactor. *ACS Catal.* **7,** 7629–7633 (2017).
18. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun.* **6,** 7789 (2015).
19. Juhász, T., Szeltner, Z. & Polgár, L. Properties of the prolyl oligopeptidase homologue from *Pyrococcus furiosus*. *FEBS Lett.* **580,** 3493–3497 (2006).
20. Polgár, L. The prolyl oligopeptidase family. *Cell. Mol. Life Sci.* **59,** 349–362 (2002).
21. Yang, H., Srivastava, P., Zhang, C. & Lewis, J. C. A general method for artificial metalloenzyme formation through strain-promoted azide–alkyne cycloaddition. *ChemBioChem* **15,** 223–227 (2014).
22. Harris, M. N., Madura, J. D., Ming, L.-J. & Harwood, V. J. Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile *Pyrococcus furiosus*. *J. Biol. Chem.* **276,** 19310–19317 (2001).
23. Wilson, Y. M., Dürrenberger, M., Nogueira, E. S. & Ward, T. R. Neutralizing the detrimental effect of glutathione on precious metal catalysts. *J. Am. Chem. Soc.* **136,** 8928–8932 (2014).
24. Reetz, M. T. *et al.* A robust protein host for anchoring chelating ligands and organocatalysts. *ChemBioChem* **9,** 552–564 (2008).
25. Drummond, D., Iverson, B., Georgiou, G. & Arnold, F. H. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.* **350,** 806–816 (2005).
26. Dąbrowski, S. & Kur, J. Cloning and expression in *Escherichia coli* of the recombinant his-tagged DNA polymerases from *Pyrococcus furiosus* and *Pyrococcus woesei*. *Protein Expr. Purif.* **14,** 131–138 (1998).
27. Punna, S., Kaltgrad, E. & Finn, M. G. 'Clickable' agarose for affinity chromatography. *Bioconj. Chem.* **16,** 1536–1541 (2005).
28. Davies, R. R. *et al.* Artificial metalloenzymes based on protein cavities: exploring the effect of altering the metal ligand attachment position by site directed mutagenesis. *Bioorg. Med. Chem. Lett.* **9,** 79–84 (1999).
29. Filice, M. *et al.* Synthesis of a heterogeneous artificial metallolipase with chimeric catalytic activity. *Chem. Commun.* **51,** 9324–9327 (2015).
30. Ball, Z. T. Designing enzyme-like catalysts: a rhodium(II) metallopeptide case study. *Acc. Chem. Res.* **46,** 560–570 (2013).
31. Renata, H. *et al.* Identification of mechanism-based inactivation in P450-catalyzed cyclopropanation facilitates engineering of improved enzymes. *J. Am. Chem. Soc.* **138,** 12527–12533 (2016).
32. Popp, B. V. & Ball, Z. T. Proximity-driven metallopeptide catalysis: remarkable side-chain scope enables modification of the Fos bZip domain. *Chem. Sci.* **2,** 690–695 (2011).
33. Lewis, J. C. & Arnold, F. H. Catalysts on demand: selective oxidations by laboratory-evolved cytochrome P450 BM3. *CHIMIA* **63,** 309–312 (2009).
34. Lu, Y., Yeung, N., Sieracki, N. & Marshall, N. M. Design of functional metalloproteins. *Nature* **460,** 855–862 (2009).
35. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
36. Kaushik, S., Etchebest, C. & Sowdhamini, R. Decoding the structural events in substrate-gating mechanism of eukaryotic prolyl oligopeptidase using normal mode analysis and molecular dynamics simulations. *Proteins* **82,** 1428–1443 (2014).

Acknowledgements

This work was supported by, or in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant nos. W911NF-14-1-0334 and 66796-LS-RIP (to J.C.L.), the National Science Foundation (NSF) under CAREER Award CHE-1351991 (to J.C.L.), The David and Lucile Packard Foundation (to J.C.L.), the National Institutes of Health (NIH) National Cancer Institute (R00CA175399 to R.E.M.), the Damon Runyon Cancer Research Foundation (DFS-08-14 to R.E.M.) and the NSF under the Center for Chemical Innovation Center for Selective C–H Functionalization (CHE-1700982, to J.C.L.). K.E.G. and D.M.U. were funded by an NIH Chemistry and Biology Interface Training Grant (T32 GM008720) and G.L. was supported by the Kwanjeong Educational Foundation. MS data were acquired on instruments purchased using an NSF instrumentation grant (CHE-1048528).

Author contributions

H.Y. and P.S. developed the soluble ArM evolution procedure. C.Z. optimized the soluble ArM evolution procedure. A.M.S. and D.M.U. optimized the soluble ArM evolution procedure and collected conversion and selectivity data for all evolved ArMs. H.J.P. optimized and executed the immobilized ArM evolution procedure. K.B. and Y.G. cloned POP amber mutants and the combinatorial codon mutagenesis library. K.E.-G., G.L. and R. E.M. conducted and analysed the LC-MS/MS experiments. J.C.L. devised the experiments and procedures, designed the ArM variants and libraries, analysed data and wrote the manuscript.

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. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.C.L.

Competing financial interests

The authors declare no competing financial interests.