



Methods for the directed evolution of proteins

Michael S. Packer and David R. Liu

Abstract | Directed evolution has proved to be an effective strategy for improving or altering the activity of biomolecules for industrial, research and therapeutic applications. The evolution of proteins in the laboratory requires methods for generating genetic diversity and for identifying protein variants with desired properties. This Review describes some of the tools used to diversify genes, as well as informative examples of screening and selection methods that identify or isolate evolved proteins. We highlight recent cases in which directed evolution generated enzymatic activities and substrate specificities not known to exist in nature.

Natural selection

A process by which individuals with the highest reproductive fitness pass on their genetic material to their offspring, thus maintaining and enriching heritable traits that are adaptive to the natural environment.

Artificial selection

(Also known as selective breeding). A process by which human intervention in the reproductive cycle imposes a selection pressure for phenotypic traits desired by the breeder.

Libraries

Diverse populations of DNA fragments that are subject to downstream screening and selection.

Over many generations, iterated mutation and natural selection during biological evolution provide solutions for challenges that organisms face in the natural world. However, the traits that result from natural selection only occasionally overlap with features of organisms and biomolecules that are sought by humans. To guide evolution to access useful phenotypes more frequently, humans for centuries have used artificial selection, beginning with the selective breeding of crops¹ and domestication of animals². More recently, directed evolution in the laboratory has proved to be a highly effective and broadly applicable framework for optimizing or altering the activities of individual genes and gene products, which are the fundamental units of biology.

Genetic diversity fuels both natural and laboratory evolution. The occurrence rate of spontaneous mutations is generally insufficient to access desired gene variants on a time scale that is practical for laboratory evolution. A number of genetic diversification techniques are therefore used to generate libraries of gene variants that accelerate the exploration of a gene's sequence space. Methods to identify and isolate library members with desired properties are a second crucial component of laboratory evolution. During organismal evolution, phenotype and genotype are intrinsically coupled within each organism. However, during laboratory evolution (FIG. 1a), it is often inconvenient or impossible to manipulate genes and gene products in a coupled manner. Therefore, single-gene evolution in the laboratory requires carefully designed strategies for screening or selecting functional variants in ways that maintain the genotype–phenotype association.

In this Review, we summarize techniques that generate single-gene libraries, including standard methods as well as novel approaches that can generate superior diversity containing a larger proportion of functional mutants. We also review screening and selection methods that identify or isolate improved variants within these libraries. Although these strategies can be applied to multi-gene pathways^{3,4} and gene networks^{5–7}, the examples in this Review will focus exclusively on the laboratory evolution of single genes. In addition, although many of these approaches apply to other types of biomolecules, we focus on the directed evolution of proteins because protein evolution has proved to be especially useful for generating novel biocatalysts⁸, reagents⁹ and therapeutics¹⁰.

Methods for gene diversification

It is impossible to cover the entire mutational space of a typical protein: complete randomization of a mere decapeptide would yield 10^{13} unique combinations of amino acids, which exceeds the achievable library size of almost all known protein library creation methods. Because comprehensive coverage of sequence space is impossible, gene diversification strategies are designed to perform an optimal sparse sampling of a vast multi-dimensional sequence space. The activity level of each library member can be conceptualized as the elevation in a fitness landscape on an x – y coordinate that represents the genotype of that library member. The goal of directed evolution studies is to take mutational steps within this landscape that ‘climb’ towards peak activity levels (FIG. 1b). Over many generations, these beneficial mutations accumulate, resulting in a successively improved phenotype.

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.

Correspondence to D.R.L.
e-mail: drlu@fas.harvard.edu
doi:10.1038/nrg3927
Published online 9 June 2015

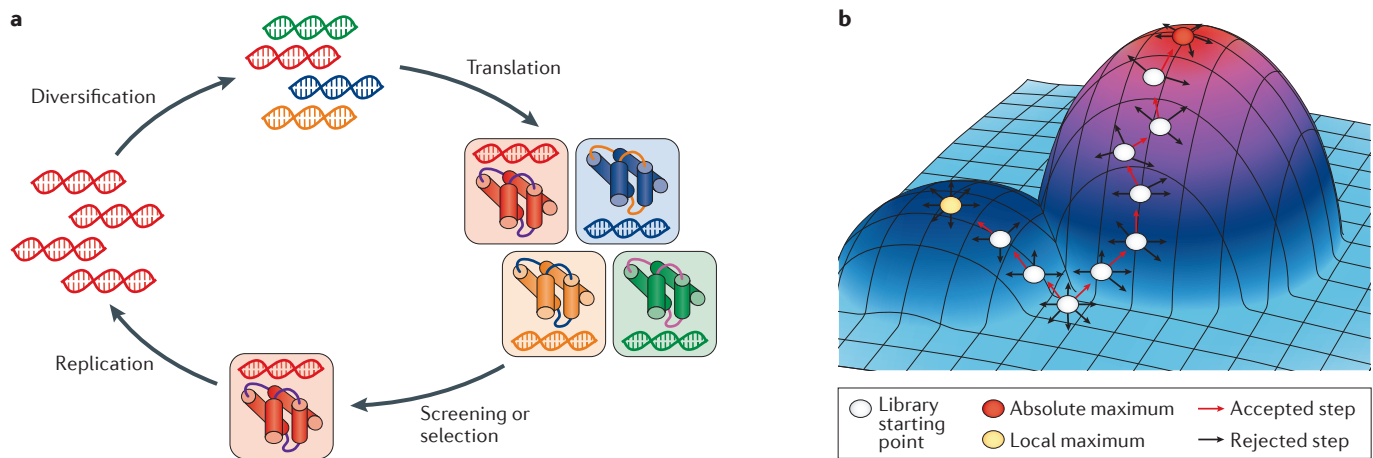


Figure 1 | Key steps in the cycle of directed evolution. **a** | The process of directed evolution in the laboratory mimics that of biological evolution. A diverse library of genes is translated into a corresponding library of gene products and screened or selected for functional variants in a manner that maintains the correspondence between genotype (genes) and phenotype (gene products and their functions). These functional genes are replicated and serve as starting points for subsequent rounds of diversification and screening or selection. **b** | Although the mutational space is multidimensional, it is conceptually helpful to visualize directed evolution as a series of steps within a three-dimensional fitness landscape. Library generation samples the proximal surface of the landscape, and screening or selection identifies the genetic means to ‘climb’ towards fitness peaks. Directed evolution can arrive at absolute maximum activity levels but can also become trapped at local fitness maxima in which library diversification is insufficient to cross ‘fitness valleys’ and access neighbouring fitness peaks.

Library size

The number variants that are subjected to screening and selection. Library sizes are limited by molecular cloning protocols and/or by host transformation efficiency.

Focused mutagenesis

A strategy of diversification that introduces mutations at DNA regions expected to influence protein activity.

Random mutagenesis

A strategy of diversification that introduces mutations in an unbiased manner throughout the entire gene.

Mutational spectrum

The frequency of each specific type of transition and transversion. The evenness of this spectrum allows more thorough sampling of sequence space.

Transformation

The process by which a cell directly acquires a foreign DNA molecule. A number of protocols allow high-efficiency transformation of microorganisms through treatments with ionic buffers, heat shock or electroporation.

Neutral drift

A process that occurs in the presence of a purifying selection pressure to eliminate deleterious mutations. This is in contrast to genetic drift, a process by which mutations fluctuate in frequency in the absence of selection pressure.

Researchers can use focused mutagenesis to maximize the likelihood that a library contains improved variants, provided that amino acid positions that are likely determinants of the desired function are known. In the absence of plausible structure–function relationships, random mutagenesis can provide a greater chance of accessing functional library members than focusing library diversity on incorrectly chosen residues that, when mutated, do not confer desired activities. Researchers have developed an extensive range of methods to perform both forms of gene diversification, and the most successful strategies often integrate random and focused mutagenesis.

Random mutagenesis. Traditional genetic screens use chemical and physical agents to randomly damage DNA. These agents include alkylating compounds such as ethyl methanesulfonate (EMS)¹¹, deaminating compounds such as nitrous acid¹², base analogues such as 2-aminopurine¹³, and ultraviolet irradiation¹⁴. Chemical mutagenesis is sufficient to deactivate genes at random for a genome-wide screen but is less commonly used for directed evolution because of biases in mutational spectrum^{11,12}.

Non-chemical methods to randomly mutate genes frequently enhance the rate of errors during DNA replication. In *Escherichia coli*, DNA replication by DNA polymerase III introduces mutations at a rate of 10⁻¹⁰ mutations per replicated base¹⁵. This rate is increased in mutator strains containing deactivated proofreading and repair enzymes, mutS, mutT and mutD¹⁵⁻¹⁷. Transformation of the XL1-red strain with a plasmid bearing the evolving gene yields mutations at a rate of 10⁻⁶ per base per generation¹⁶. Unfortunately, these strains

not only mutate the library member but also induce deleterious mutations in the host genome. Host intolerance to a high degree of genomic mutation places an upper limit on *in vivo* mutagenesis rates. To avoid this constraint, C. C. Liu and co-workers¹⁸ developed orthogonal *in vivo* DNA replication machinery that only mutates target DNA. This method co-opts naturally occurring *Kluyveromyces lactis* linear plasmids pGKL1/2 and their specialized TP-DNA polymerases. Because this plasmid is exclusively cytoplasmic, the TP-DNA polymerase exerts no mutational load on the host genome within the nucleus of *Saccharomyces cerevisiae*.

The relatively low mutation rates and the lack of control offered by most previously described *in vivo* random mutagenesis protocols have led to a strong preference towards *in vitro* random mutagenesis strategies. In error-prone PCR (epPCR), first described by Goeddel and co-workers¹⁹, the low fidelity of DNA polymerases under certain conditions generates point mutations during PCR amplification of a gene of interest. Increased magnesium concentrations, supplementation with manganese or the use of mutagenic dNTP analogues²⁰ can reduce the base-pairing fidelity and increase mutation rates to 10⁻⁴–10⁻³ per replicated base²¹. Because mutations during PCR accumulate with each cycle of amplification, it is possible to increase the average number of mutations per clone by increasing the number of cycles.

One application of epPCR is to generate neutral drift libraries. Before directed evolution experiments are carried out, a target gene is mutagenized by epPCR and fused to a GFP reporter, and the variants are then screened for proper protein expression²². After multiple rounds of mutagenesis and screening, the resulting neutral drift library exhibits sequence diversity that does

not destabilize protein structure and is therefore largely devoid of the deleterious mutations that would otherwise have accumulated during the multiple rounds of mutagenesis. Such libraries provide a valuable and evolvable starting point for subsequent directed evolution of the target protein towards a phenotype of interest²².

The DNA polymerases used in epPCR exhibit mutational biases, but unbalanced dNTP concentrations and proprietary mixtures of polymerases can help to reduce imbalance in the mutational spectrum^{23,24}. To yield a more ideal nucleotide mutational spectrum, Schwaneberg and co-workers²⁵ developed sequence saturation mutagenesis (SeSaM) in which the universal base deoxyinosine is enzymatically inserted throughout the target gene. Although this approach is effective, epPCR is easier to implement and can provide high mutation rates with fairly broad mutational spectra.

Focused mutagenesis strategies. Many proteins are structurally characterized at sufficient resolution to implicate specific residues in substrate binding or catalysis. Although random mutagenesis can generate stochastic point mutations at codons corresponding to these residues, access to codons that require mutation of more than one nucleotide relative to the initial codon often requires a focused mutagenesis strategy. Perhaps the most straightforward focused mutagenesis approach uses synthetic DNA oligonucleotides containing one or more degenerate codons at positions corresponding to targeted residues. This mutagenic oligonucleotide is incorporated into a gene library as a mutagenic cassette²⁶ using either traditional restriction enzyme cloning or contemporary gene assembly protocols^{27–29}. The simultaneous saturation mutagenesis of multiple residues can access combinations of mutations that may exhibit epistatic interactions. For example, synergistic mutations are those that in combination confer an effect that is larger than the sum of the effects of each individual mutation. Two beneficial mutations that exhibit synergism can undergo sequential enrichment and are therefore accessible through iterative single-residue saturation libraries. However, to access combinations of mutations exhibiting sign epistasis — a case in which mutations may be deleterious in isolation but confer gain of function in combination — sequential acquisition is impossible, and simultaneous saturation is therefore necessary.

As the number of unique sequences increases exponentially with the number of randomized sites, only a handful of residues can be randomized if complete coverage of the resulting combinations of mutations is desired. Furthermore, the vast majority of individual mutations are likely to be neutral or deleterious to the desired activity³⁰. The mutational load of simultaneous saturation increases with the number of randomized sites, and the resulting library will be populated with a larger fraction of inactive library members.

For this reason, a number of focused mutagenesis strategies only introduce specific amino acid substitutions that are likely to be beneficial. Phylogenetic analyses of homologous proteins, which are pre-enriched for functional variation owing to natural selection, are one

means for identifying these potentially beneficial mutations. Wyss and co-workers³¹ demonstrated that the introduction of consensus mutations can improve thermostability and native enzymatic activity. Rather than focusing on common ancestral mutations, reconstructed evolutionary adaptive path (REAP) analysis identifies significant mutational divergence that is more likely to confer novel gain of function³². These mutational signatures can be adopted from a distinct evolutionary pathway with known phenotypic characteristics and further curated based on structural proximity to the active site.

Molecular modelling can also predict specific amino acid substitutions that are likely to be beneficial³³. Algorithms such as Rosetta calculate free energies based on steric clashes, hydrophobic packing, hydrogen bonding and electrostatic interactions³⁴. Mutations that are predicted to stabilize protein folding³⁵ or to improve transition state stabilization can be introduced into the library semi-stochastically by incorporating synthetic oligonucleotides via gene reassembly (ISOR)³⁶.

Diversification by recombination. The reassortment of mutations to access beneficial combinations of mutations is a crucial component of biological evolution. This natural process can be mimicked by a variety of methods under the broad umbrella of homologous recombination. The original DNA shuffling method described by Stemmer³⁷ fragments a gene with DNase and then allows fragments to randomly prime one another in a PCR reaction without added primers (FIG. 2a). A related method developed by Monticello and colleagues³⁸, random chimeragenesis on transient templates (RACHITT), also uses DNase-mediated fragmentation but a different method of reassembly. Fragments anneal directly to a temporary uracil-containing scaffold; upon flap resection and fragment ligation, the scaffold is digested. DNase concentration and fragmentation reaction duration offer crude mechanisms to shift fragment sizes and crossover frequencies, but newer protocols provide greater control. Nucleotide exchange and excision technology (NEXt)³⁹ incorporates a fixed concentration of deoxyuridine triphosphate (dUTP) during PCR; subsequent treatment with uracil deglycosylases and apurinic/apyrimidinic lyases yields random fragments with size distribution determined by dUTP concentration. Unlike fragmentation-based methods, staggered extension process (StEP) described by Arnold and colleagues⁴⁰ is a modified PCR protocol in which the elongation step is interrupted prematurely by heat denaturation. Subsequent annealing allows incomplete extension products to switch templates, effecting recombination of multiple DNA templates into one amplicon (FIG. 2b).

With the decreasing cost of synthetic oligonucleotides, assembly PCR⁴¹ (also known as assembly of designed oligonucleotides (ADO) or synthetic shuffling^{42,43}) has become a preferred recombination strategy. In these reactions, overlapping primers extend one another; after multiple cycles the process yields full-length gene products in which each combination of mutation-bearing oligonucleotides has been recombined (FIG. 2c).

Degenerate codons

Codons constructed with a mixed population of nucleotides at a given position, thus sampling all possible amino acids within the constructed libraries. The most popular examples are NNK and NNS (where N can be any of the four nucleotides, K can be G or T, and S can be G or C).

Epistatic interactions

Non-additive effects between mutations (for example, mutational synergy or synthetic lethality). As a result, the sequential acquisition of mutations is not always equivalent to mutational co-occurrence.

Homologous recombination

A process by which separate pieces of DNA swap genetic material, guided by the annealing of complementary DNA fragments.

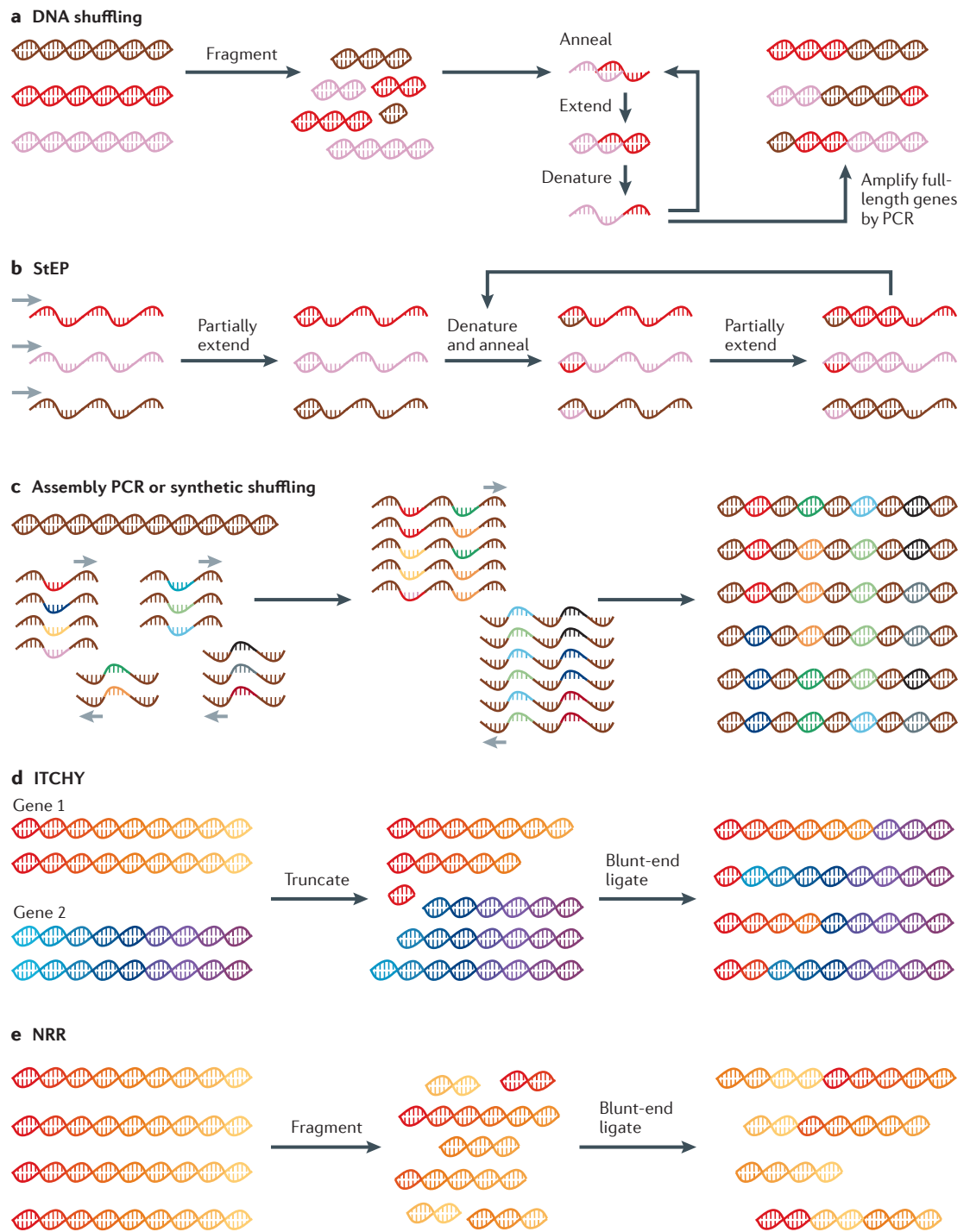


Figure 2 | **Genetic recombination methods for protein evolution.** **a** | DNA shuffling begins with DNase fragmentation of homologous gene variants. These fragments prime one another in a PCR reaction. The cycle of annealing, extending and denaturing fragments reassembles full-length gene products containing recombined segments. **b** | Staggered extension process (StEP) achieves homologous recombination by repeated premature denaturation during the extension stage of a PCR. Because partially polymerized genes may switch templates, the resulting full-length gene products are chimeras with a varying number of crossovers. **c** | Assembly PCR or synthetic shuffling methods require overlapping libraries of oligonucleotides that encode genetic variation at a number of loci. These primers extend one another in various combinations to yield recombined gene products. **d** | Incremental truncation for the creation of hybrid enzymes (ITCHY) uses exonucleases to degrade the parts of genes that encode the free amino or carboxyl termini of prospective fusion protein partners. Blunt-end ligation yields single crossovers with varying fractional composition from each template gene. **e** | Non-homologous random recombination (NRR) uses DNase fragmentation followed by blunt-end ligation to generate diverse topological rearrangements (deletions, insertions and domain reordering). Note that DNA fragments are not drawn to scale. Typical recombined fragments range in length from dozens to hundreds of bases.

Recombination is only effective on a diverse population of functional genes. Typically, one of the recombination methods described above is used between rounds of evolution to recombine mutations from distinct clones⁴⁴. Alternatively, homologous recombination with copies of the wild-type DNA sequence can eliminate non-beneficial passenger mutations, analogous to the traditional breeding technique of back-crossing³⁷. In another effective use of homologous recombination during gene diversification, a family of closely related naturally occurring homologues can be shuffled into a starting library to take advantage of nature's pre-evolved repertoire of functional gene variants⁴⁵.

Methods for *in vitro* recombination require substantial manual manipulation and are usually followed by transformation or transduction to introduce the recombined gene population back into cells. Cornish and colleagues⁴⁶ harnessed the power of native systems in *S. cerevisiae* to perform homologous recombination between a library of donor cassettes and the evolving gene. Through yeast mating, functional gene variants undergo reassortment with different donor cassettes, allowing homologous recombination within the evolving population. Seamless alternation between sexual reproduction and selection can support continuous evolution⁴⁶.

The recombination methods described above rely on sequence homology to preserve gene structure among recombinants. By contrast, sequence homology-independent protein recombination (SHIPREC) permits shuffling of disparate gene elements. Such a capability is particularly useful for recombining families of proteins with similar functions but disparate sequences⁴⁷. Homology-independent recombination can also create combinatorial protein libraries that do not preserve the ordering or lengths of domains. Ostermeier *et al.*⁴⁸ devised incremental truncation for the creation of hybrid enzymes (ITCHY), in which homology-independent recombination is used to create hybrid enzymes through the incremental truncation and fusion of two distinct genes (FIG. 2d). In addition, our laboratory⁴⁹ used non-homologous random recombination (NRR) to generate functional proteins with substantial rearrangements of domain topology (FIG. 2e). Although these two techniques are different, they both involve random fragmentation (for example, using a DNase or an exonuclease) followed by sequence-independent ligation of fragments. Tuning fragmentation conditions can shift the average number of crossovers, and electrophoresis can be used to isolate ligated products of the desired length to minimize inactive library members that are too short or too long, or that have excessive numbers of crossovers.

Nonetheless, the vast majority of non-homologous recombinants will display domain disruption and folding instability. The SCHEMA algorithm computationally identifies breakpoints in proteins that minimize the number of inter-domain interactions⁵⁰. Type IIb restriction enzyme sites can be inserted at these optimal breakpoints within the DNA sequences, and enzymatic digestion yields 'sticky ends' that enable sequence-independent site-directed chimeragenesis (SISDC)⁵¹.

Alternatively, chimeric oligonucleotides with complementarity to two distinct domains defined either by eukaryotic exons⁵² or by SCHEMA can be used in overlap extension PCR⁵³. A library of these chimeric primers can be used to shuffle domains even in the complete absence of homology.

Diversification strategy considerations. Directed evolution practitioners increasingly use sophisticated focused mutagenesis methods to construct smaller libraries of higher quality that sample a functionally rich portion of the fitness landscape. These strategies require phylogenetic information or molecular structures to focus library diversity on residues or even specific substitutions that are thought to be necessary for the desired activity. In the absence of this information, random mutagenesis is an absolute necessity. Even when the requisite data are available, deducing the determinants of protein function at the amino acid level can be challenging. Random mutagenesis maybe used to probe mutations that are distant from obvious substrate contact sites or that are not present in naturally evolved orthologues. Fortunately, random and focused mutagenesis strategies can be combined into a single diversification step or applied separately during successive rounds of evolution to maximize the likelihood of success⁵⁴ (TABLE 1).

Genetic screens for single-gene evolution

Genetic screens were originally developed to discover genes associated with specific phenotypes. Geneticists randomly mutagenize the genome of a model organism and then assay individual organisms for a phenotype of interest. Organisms with altered phenotypes are characterized by crossing and linkage analyses, or more recently by high-throughput DNA sequencing, to identify specific mutations underlying phenotypic changes. Directed evolution applies similar screening strategies to single-gene libraries prepared with the aforementioned diversification methods.

Screens of spatially separated variants. Spatial separation (that is, encoding by location) of individual mutants preserves the linkage between phenotype and genotype. For these screens, gene variants are expressed in a unicellular model organism such as *E. coli* that can be screened as colonies on solid media or transferred into multiwell liquid culture plates (FIG. 3A). Although spatial separation of clones imposes a practical throughput limit of fewer than $\sim 10^4$ library members per screening round, a key advantage of this approach is its broad compatibility with many different assay techniques. When a fluorescent readout is not available, techniques such as nuclear magnetic resonance (NMR), high-performance liquid chromatography (HPLC), gas chromatography or mass spectroscopy can directly monitor substrate consumption or product formation. In principle, almost any enzymatic activity can be screened in a spatially separated library format, although the time-consuming and infrastructure-intensive nature of some spatially separated screening techniques further limit throughput.

Passenger mutations

(Also known as hitchhiker mutations). Unnecessary mutations that are enriched in a population owing to co-occurrence with a highly beneficial linked mutation.

Transduction

The process by which a viral vector delivers a foreign DNA molecule to a cellular host.

Table 1 | Comparison and summary of approaches to library diversification

Category	Examples and methods	Random or focused?	In vivo or in vitro?	Advantages	Disadvantages
Chemical mutagenesis ^{11,12}	EMS, nitrous acid, ultraviolet irradiation and bisulfite	Random	<i>In vitro</i> and <i>in vivo</i>	Dose-dependent mutation rates	Low mutation rates; uneven mutational spectrum; hazardous chemicals
Mutator strains	XL1-red <i>E. coli</i> ¹⁶ , mutagenesis plasmid (PACE) ¹²² and yeast orthogonal replication ¹⁸	Random	<i>In vivo</i>	Easy to use	Low mutation rates; uneven mutational spectrum
epPCR	Taq supplemented with Mg ²⁺ , Mn ²⁺ and/or unequal dNTPs ²³ ; proprietary enzyme mixes (Mutazyme) ²⁴	Random	<i>In vitro</i>	Permits high mutation rates; easy to use commercial formulations; relatively even mutational spectrum	Random mutagenesis at the nucleotide level but does not evenly sample amino acid codon space
Site-directed saturation mutagenesis	NNK and NNS codons (where N can be any of the four nucleotides, K can be G or T, and S can be G or C) on mutagenic primers ³⁰	Focused	<i>In vitro</i>	Fully samples amino acid repertoire; focus on functionally relevant residues increases library quality	Requires structural or biochemical knowledge; excess of inactive clones within simultaneous saturation libraries
Computational strategies for high-quality library design	Rosetta design and computationally guided libraries ³⁵ , ISOR ³⁶ , consensus design ³¹ , REAP ³² and SCHEMA ⁵⁰	Focused	<i>In vitro</i>	Can create small libraries pre-enriched for functional variation by natural selection and/or <i>in silico</i> filtering	Requires structural data or phylogenetic data
Homologous recombination	DNA shuffling ³⁷ , family shuffling ⁴⁵ , StEP ⁴⁰ , RACHITT ³⁸ , NEX ³⁹ , heritable recombination ⁴⁶ , ADO ⁴³ and synthetic shuffling ⁴²	NA	<i>In vitro</i> or <i>in vivo</i>	Can identify beneficial combinations of mutations or eliminate passenger mutations; can also shuffle sequences of orthologous proteins to repurpose functional diversity from nature	Rely heavily on sequence homology; evolved clones and natural orthologues can be divergent in nucleotide sequence
Non-homologous recombination	ITCHY ⁴⁸ , SHIPREC ⁴⁷ , NRR ⁴⁹ , SISDC ⁵¹ and overlap extension PCR ^{52,53}	NA	<i>In vitro</i>	Capable of shuffling distantly related sequences, permuting order and combinatorial gene fusions	Highly effective in niche applications but challenging to implement as a general strategy

ADO, assembly of designed oligonucleotides; *E. coli*, *Escherichia coli*; EMS, ethyl methanesulfonate; epPCR, error-prone PCR; ISOR, incorporating synthetic oligonucleotides via gene reassembly; ITCHY, incremental truncation for the creation of hybrid enzymes; NA, not applicable; NEX, nucleotide exchange and excision technology; NRR, non-homologous random recombination; PACE, phage-assisted continuous evolution; RACHITT, random chimeragenesis on transient templates; REAP, reconstructed evolutionary adaptive path; SHIPREC, sequence homology-independent protein recombination; SISDC, sequence-independent site-directed chimeragenesis; StEP, staggered extension process.

When performing low-throughput screens, an understanding of structure–activity relationships within the target protein may be necessary to maximize the probability of accessing a desired variant. These considerations are best exemplified by the evolution of cytochromes P450, a class of enzymes with high evolutionary potential evidenced by the diverse oxidative reactions they catalyse in nature. Arnold and colleagues⁸ screened a panel of ~100 previously designed P450 variants in *E. coli* lysates for carbene transfer to form cyclopropanes; product formation was monitored by gas chromatography. The resulting enzymes exhibit high-activity cyclopropanation with enantioselectivity and diastereoselectivity, capabilities that are not known to exist in any natural biocatalysts. In this case, prior knowledge of mutants with altered P450 activities enabled success with only a small library and a low-throughput screen.

When molecular insight or prior knowledge is lacking, it may be necessary to screen more variants to reach the desired phenotype. High-throughput screens rely on the rapid assessment of optical features such as colour, fluorescence, luminescence or turbidity. In special cases, the protein of interest has an inherently visible

phenotype, as demonstrated by the pioneering evolution of the alkaline serine protease subtilisin. You and Arnold⁵⁵ screened colonies on casein plates for zones of clearing due to proteolysis of substrate milk proteins. A secondary screen on casein plates containing dimethylformamide (DMF) identified variants exhibiting solvent-tolerant proteolysis.

Fluorescent proteins provide a readily screenable phenotype, and thus multiple research groups have used cellular fluorescence as a screen to identify GFP variants with brighter fluorescence and altered absorption or emission spectra^{44,56}. More recently, this approach was applied to Arch rhodopsin, a form of channel rhodopsin engineered by Cohen and colleagues⁵⁷ to exhibit voltage-dependent fluorescence and used to directly image neuronal activity. Arnold and colleagues⁹ expressed a library of Arch variants in *E. coli* using multiwell liquid culture plates and washed cells with ionic buffer to generate the transmembrane potential required for fluorescence measurements. After multiple rounds of screening random and site-directed libraries, the most active variant displayed red-shifted emission and increased brightness. The capabilities of evolved Arch

Evolutionary potential

The capacity of a protein to take on new functions through evolution. High thermostability allows for necessary but destabilizing mutations, and functional diversity of homologues is a demonstration of previous evolution in nature.

Surrogate substrates
Substrate analogues that are permissive of enzymatic conversion but that, upon catalysis, exhibit chemical rearrangements that lead to altered optical properties, including visible colour, relief of fluorophore quenching, shifted fluorophore excitation or emission, and downstream chemiluminescence.

should enable parallel monitoring of multiple neurons using wide-field microscopy.

Most biomolecules are not associated with directly observable phenotypes and therefore require a fluorescent, colorimetric or other readily detectable reporter. Surrogate substrates can be added directly to liquid culture or lysates to generate a fluorescent, luminescent or colorimetric signal that is proportional to the enzymatic activity of interest. As a result, these reporters allow precise screening of diverse catalysts such as P450 monooxygenases⁵⁸, cellulases⁵⁹, organophosphate hydrolases⁶⁰ and retroaldolases⁶¹. However, the development of surrogate substrates for some reactions can represent a substantial undertaking⁶². In addition, evolved variants

will have only been screened for activity on a surrogate substrate, and they must be separately assayed to ensure that enzyme optimization on the surrogate also improves activity on the desired substrate.

Widely used genetic reporters such as *GFP*, luciferase and beta-galactosidase enable facile detection of gene expression. Expression-mediated screens have been developed for the study of protein–protein interactions⁶³ and the activity of enzymes including cellulases and glycosynthases^{64–66}. As a general strategy, small-molecule- or cell-state-inducible genetic circuitry from nature can be used to detect desired enzymatic activity. For example, Ackerley and co-workers⁶⁷ used the DNA-damage inducible SOS promoter to express

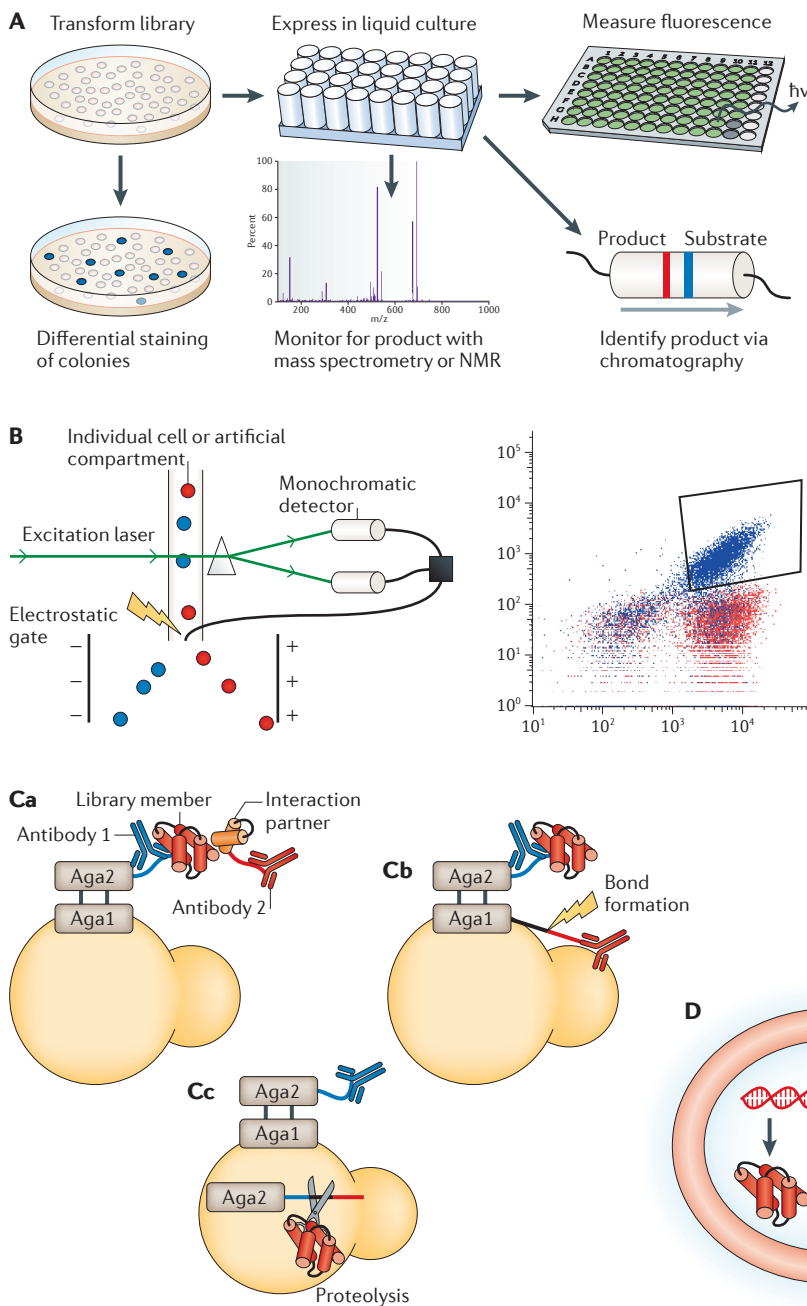


Figure 3 | Screening methods for protein evolution.

A | Clonally isolated variants are screened as colonies on solid media or as wells in liquid culture. Fluorescent or colorimetric reporters are measured by automated microtitre plate readers. Alternatively, lysates can be screened for product formation using chromatography, mass spectrometry or nuclear magnetic resonance (NMR). **B** | Fluorescence-activated cell sorting (FACS) enables the fluorescence measurement of individual cells and the separation of distinct subpopulations by electrostatic deflection. **C** | Yeast display techniques enable FACS screens of protein–protein interactions (part **Ca**), bond formation (part **Cb**) and peptide bond cleavage (part **Cc**). An evolving protein is displayed on the cell surface as a fusion to the cell–cell adhesion protein Aga2. Novel protein–protein interactions or improved affinities are identified on the basis of decoration with fluorescent-labelled antibodies that recognize both the evolving protein and a bound target (part **Ca**). Similarly, surface-displayed protein ligases that have high activity or altered substrate specificity are sorted on the basis of the covalent attachment of an epitope tag that is then detected by a fluorescent-labelled antibody (part **Cb**). In a screen for proteases with altered substrate specificities, the enzyme is retained in the endoplasmic reticulum. Upon substrate cleavage, an epitope tag is removed from an Aga2 fusion protein, thus allowing differential labelling with antibodies (part **Cc**). **D–E** | *In vitro* compartments can be formed through double emulsions (part **D**) or with polyelectrolyte shells (part **E**). These compartments entrap DNA, translated proteins and fluorogenic substrates, allowing the fluorescence-activated sorting of functional variants. m/z, mass-to-charge ratio.

beta-galactosidase in proportion to nitroreductase activation of genotoxic prodrugs. Through iterative site-directed mutagenesis, this screen identified nitroreductase variants that activated chemotherapeutic prodrugs and killed tumour cells with greater efficiency than wild-type nitroreductase. Gene expression reporters are imperfect measures of enzymatic activity but, when used properly, can correlate strongly with enzymatic activity⁶⁸. Automated fluorescence measurement and robotic colony picking lighten the tedious workload of these screens, but the physical and material constraints associated with spatial separation inherently limit throughput.

High-throughput screening by flow cytometry. Rather than spatially separating clones, a bulk population can be interrogated at the level of individual cells using the cell wall or membrane to maintain genotype–phenotype association. Fluorescence-activated cell sorting (FACS)⁶⁹ relies on a non-diffusing fluorescent reporter to automate the identification and isolation of cells containing desired gene variants (FIG. 3B). Integrating major advances in microfluidics, optics and cell manipulation, state-of-the-art flow cytometry offers one of the highest capacities of any screening method, achieving up to 10⁸ library members screened in <24 hours^{70,71}.

Cytosolic fluorescent or luminescent proteins within cells can form the basis for FACS screens of enzymes such as recombinases, chaperones and inteins^{72–74}. Cell surface-displayed epitopes are also non-diffusive and can be detected by FACS using fluorescent-labelled antibodies. This approach became more widely used with the development of a yeast display screen for protein–protein interactions⁷¹. Boder and Wittrup⁷¹ expressed a library of epitope-tagged antibody fragments fused to the yeast mating adhesion receptor Aga2. The resulting library members were displayed on the surface of cells, where they had the opportunity to bind to a target protein fused to a second epitope tag. FACS enabled the isolation of cells decorated with two fluorescent-labelled antibodies, one for each of the epitopes, indicating proper antibody display and target binding (FIG. 3Ca). Researchers can modulate the stringency of FACS screens by varying washing conditions and the fluorescence threshold that triggers cell isolation. For many years, yeast surface display has facilitated affinity maturation of antibody–antigen pairs⁷⁵ and the discovery of new protein–protein interactions⁷⁶.

Recently, the yeast display framework has been applied to the evolution of more diverse enzymatic activities. Bond-forming enzymes can be evolved using yeast display, as our laboratory⁷⁷ demonstrated by evolving sortase A (SrtA), a sequence-specific transpeptidase (that is, protein ligase) from *Staphylococcus aureus*. Aga2–SrtA library members were displayed on the cell surface alongside a triglycine (GGG) acceptor peptide fused to Aga1. Upon incubation with the biotinylated substrate peptide LPETG, active SrtA catalysed bond formation between the substrate and the acceptor. FACS was used to isolate cells displaying the biotinylated LPETGGG product (FIG. 3Cb). Owing to the unfavourable kinetics of wild-type SrtA, efficient bioconjugation

typically requires equimolar concentration of substrate and enzyme. Iterated rounds of FACS screening with increasing stringency produced evolved variants of SrtA (eSrtA) with 140-fold higher $k_{\text{cat}}/K_{\text{m}}$ values, enabling new applications^{78–83}.

The development of a negative screen (also known as counterscreen) using unlabelled competitor substrates enabled our laboratory⁸⁴ to evolve reprogrammed orthogonal sortases that selectively conjugate LAETG or LPESG substrates. Because substrates are applied *ex vivo*, this approach is not limited to genetically encoded peptide substrates, and it should be possible to design similar screens for enzymes that catalyse many different classes of bond-forming reactions.

Yeast display can also be modified for the evolution of bond-cleaving enzymes. Iverson, Georgiou and colleagues⁸⁵ developed yeast endoplasmic reticulum sequestration screening (YESS) in which Aga2 is expressed as a fusion protein to a negative screening substrate, epitope tag 1, a positive screening substrate and epitope tag 2. The Aga2 substrate is retained in the endoplasmic reticulum for processing by a member of a protease library. The presence of both epitope tags on the cell surface indicates protease inactivity, whereas proteolysis of the negative screening substrate would eliminate both tags. FACS isolated the subpopulation of proteases that exclusively cleaved the positive screening substrate and thereby left only epitope tag 1 on the cell surface (FIG. 3Cc). Using YESS, Iverson, Georgiou and colleagues⁸⁵ evolved tobacco etch virus (TEV) protease variants that selectively cleave ENLYFE/S or ENLYFH/S sequences but not the wild-type substrate ENLYFQ/S. These recent advances demonstrate how cell surface display can be adapted to screen for complex enzymatic activities.

Screening artificial cell-like compartments. When cell-constrained fluorescent reporters are difficult or impossible to implement for a given gene and phenotype, *in vitro* compartmentalization (IVC) provides an alternative format to enable high-throughput screening. IVC, pioneered by Tawfik and Griffiths⁸⁶, uses the aqueous droplets in water–oil emulsions to compartmentalize individual genes and gene products along with a surrogate fluorogenic substrate. IVC can enable protein evolution in two formats: either emulsion of single cells expressing the library member or emulsion of individual DNA molecules together with *in vitro* transcription–translation machinery. Because flow cytometers can only sort particles in an aqueous mixture, a secondary emulsion is necessary to create water–oil–water droplets⁸⁷ (FIG. 3D) for FACS-based screening. The flexibility to use fluorogenic substrates expands the phenotypes and enzymes that can be screened by flow cytometry.

Recently, IVC coupled with flow cytometry was used to evolve mammalian paraoxonase 1 (PON1). Wild-type PON1 can degrade a variety of organophosphate compounds and has a weak activity on some nerve agents. Tawfik and colleagues⁶⁰ used fluorogenic coumarin substrate analogues to sort IVC droplets based on phosphotriesterase activities of PON1 variants. The resulting evolved enzyme rePON1 exhibits a 10⁵-fold increase in

Fluorescence-activated cell sorting

(FACS). A flow cytometry method in which an aqueous suspension of cells or cell-like compartments is measured for fluorescence (often at multiple wavelengths) one cell at a time and subsequently separated based on a fluorescence threshold.

Negative screen

A screening method that involves depletion of an undesired phenotype.

Positive screening

Enrichment for a desired activity such as improved kinetics, tolerance to unnatural conditions and acceptance of new substrates.

catalytic activity on cyclosarin and is the first enzyme to degrade G-type (sarin-like) nerve agents with sufficient efficiency to provide prophylactic protection.

Chip-based microfluidic systems ('FACS on a chip') offer several advantages over conventional flow cytometry apparatus. The process of microfluidic droplet formation is more likely to encapsulate single cells or DNA library members, and the consistent volume and quantity of fluorescent reporters in each droplet can support highly quantitative measurements⁸⁸. Furthermore, the path length of the flow cell precisely dictates the reaction time. These advantages have been demonstrated in proof-of-concept screens for cellulase and peroxidase activities^{59,88}.

Alternative cell-like compartments beyond water-oil emulsions can also entrap genes, proteins and substrates in a suitable format for screening. Shell-like compartments made of layered polycationic and polyanionic polymers (polyelectrolytes) can encapsulate *E. coli* cells. Because these compartments are stable to detergent, DNA and protein remain linked even after detergent-induced cell lysis. Scott and Plückthun⁸⁹ used this platform to screen for properly solubilized G protein-coupled receptors (GPCRs) that retain their structure and affinity for a fluorescent probe. In a similar approach, Hollfelder and co-workers⁶⁰ built polyelectrolyte gel-shell beads (GSBs) that are compatible with flow cytometry (FIG. 3E). Using the fluorogenic organophosphate analogues described above, the researchers sorted GSBs based on phosphotriesterase activity to identify parathion hydrolase variants that more rapidly degrade organophosphate pesticides⁹⁰.

Selections for functional proteins

Screening, by definition, requires the inspection of individual phenotypes. The resulting data, which can be very rich depending on the choice of observables, not only identify desirable subpopulations but also inform the choice of appropriate screen stringency in subsequent rounds of evolution. By contrast, selection bypasses the need to individually inspect each library member and instead links an activity of interest to physical separation of the encoding DNA or to survival of the organism producing active library members. The development of effective schemes by which molecular activities of interest lead to segregation or replication of desired variants can be a major undertaking that requires creativity and strong molecular intuition. Well-designed selection offers unparalleled throughput albeit at the expense of potentially rich screening data. This drawback often necessitates a secondary phenotypic assay of selection hits in order to optimize diversification and selection protocols for the next cycle of evolution.

Selections for binding affinity. Because all library members in the same mixture undergo selection simultaneously, a molecular linkage between genes and the corresponding gene products, rather than spatial encoding, must be maintained. In a typical target-binding selection, protein library members with desired binding activity and their encoding DNA sequences are captured

using an immobilized target, whereas non-binding library members are washed away. In cell surface display or phage display methods, a cell or bacteriophage serves as a compartment to link genes and gene products. Protein library members are expressed on the surface of the cell or the coat of the bacteriophage through fusion with endogenous cell surface proteins⁹¹ or phage coat proteins^{71,92}. Phage display has proved to be highly effective in the development of therapeutic antibodies^{10,93} and in the elucidation of peptide binding motifs⁹⁴.

Unlike screening methods that are typically limited by measurement throughput, a transformation bottleneck^{95,96} restricts library sizes that can be processed by selection methods such as cell surface display or phage display, both of which require intracellular translation. As bacterial transformation provides, at best, $\sim 10^9$ – 10^{10} transformants per experiment, cell- or phage-based selection methods are generally limited to library sizes in this range⁹⁶. Ribosome display, developed by Hanes and Plückthun⁹⁷, can bypass this bottleneck through the use of *in vitro* translation reactions. In the absence of a stop codon and under carefully controlled conditions, ribosomes remain stably bound to both the mRNA and the growing polypeptide, thereby coupling proteins with their encoding genes. Similarly, mRNA display, developed by Wilson, Keefe and Szostak⁹⁸, covalently links a translated protein to its encoding mRNA through a puromycin analogue. Binding selections are conceptually simple but limited in scope (FIG. 4a). They are well suited for evolving binding affinity but have only been used in a limited number of cases to evolve enzymes, including β -lactamases⁹⁹ and RNA ligases¹⁰⁰. Although binding affinity is an important component of enzymatic activity, catalytic efficiency and the rate of product release — two properties that are not necessarily maintained or improved during a binding selection — can strongly determine overall enzyme desirability.

Organismal survival as a basis for selection. In a second important class of selections, active library members enable organisms containing their corresponding genes to survive and replicate. Antibiotic resistance is perhaps the most straightforward activity to evolve using the selective replication of *E. coli*. Numerous studies have evolved enzymes that neutralize or export antibiotics, yielding variant enzymes that are predictive of natural evolutionary trajectories in microorganisms with tolerance to higher doses of antibiotics or resistance to a broader scope of antibiotic substrates^{45,101,102}. In addition to evolving the genes that confer antibiotic resistance, it is also possible to use antibiotic selections to evolve other proteins by linking the desired activity to the expression of an antibiotic resistance gene. For example, Schultz and co-workers^{103,104} evolved aminoacyl tRNA synthetases that aminoacylate suppressor tRNAs with non-canonical amino acids, resulting in the suppression of a stop codon within a chloramphenicol efflux pump gene. In a similar strategy linking enzymatic activity to antibiotic resistance, Barbas and colleagues¹⁰⁵ evolved recombinases with altered DNA sequence specificities by using their activity to reassemble a beta-lactamase gene.

Transformation bottleneck

The efficiency at which DNA library members are transferred into the host organism, thus restricting the number of variants that can be assessed by *in vivo* selection and screening.

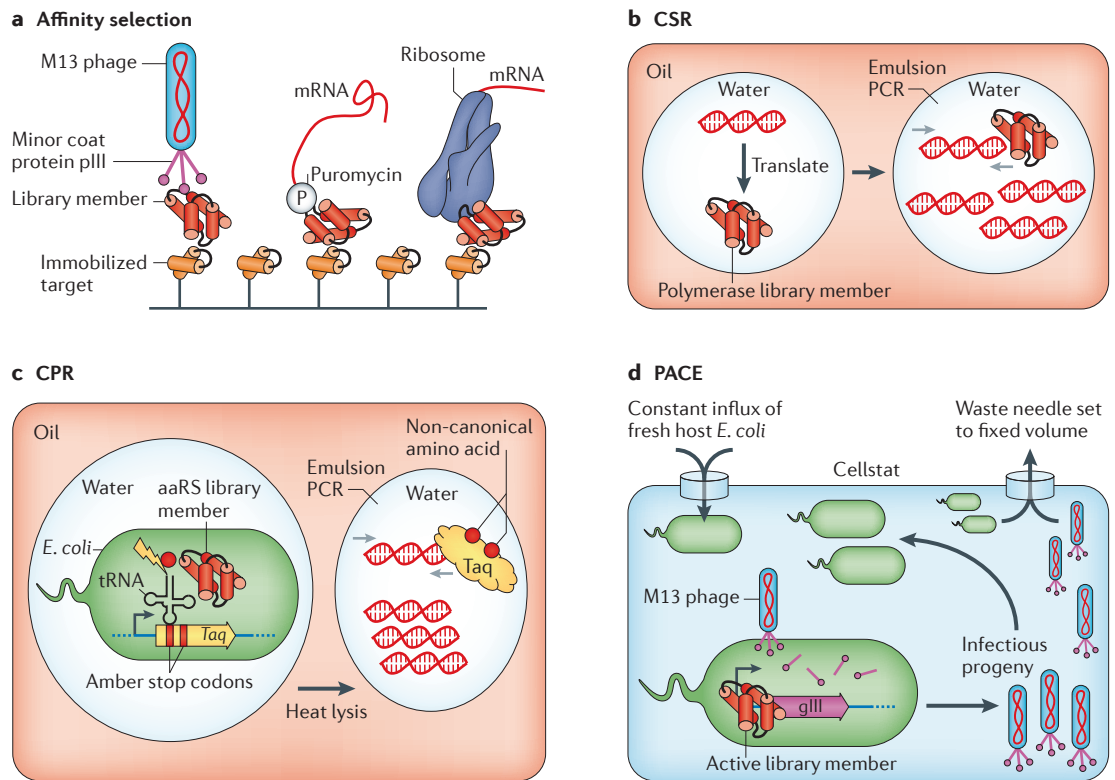


Figure 4 | Selection methods for protein evolution. **a** | Affinity selection identifies library members that bind to an immobilized target. Methods for covalently linking proteins with their corresponding genes during selection include display on phage particles via protein fusion to the coat protein pIII (left), covalent attachment to their encoding mRNA transcript via a puromycin linkage (middle) and the non-covalent attachment of both mRNA and nascent polypeptide to stalled ribosomes (right). **b** | Compartmentalized self-replication (CSR) selects for DNA and RNA polymerases that can amplify, by PCR, their own genes within water emulsion droplets (blue circle) isolated from one another by an oil phase (brown rectangle). **c** | In compartmentalized partnered replication (CPR), the evolving activity must trigger expression of Taq polymerase. For example, aminoacyl tRNA synthetase (aaRS) activity promotes amber stop codon suppression, leading to the expression of full-length Taq polymerase. Individual *Escherichia coli* cells are then isolated in water–oil emulsion droplets and lysed by heat. Higher Taq expression leads to better PCR amplification of the active library members. **d** | During phage-assisted continuous evolution (PACE), host *E. coli* cells continuously dilute an evolving population of $\sim 10^{10}$ filamentous bacteriophages in a fixed-volume vessel (cell stat; blue rectangle). Phage encoding active variants trigger host cell expression of the missing phage protein (pIII) in proportion to the desired activity and consequently produce infectious progeny, whereas phage with inactive variants produce progeny that are not infectious and are diluted out of the vessel.

Auxotroph complementation can also form the basis of selections for the evolution of metabolic enzymes. Xylose metabolism is an important target for protein evolution because xylose is a limiting factor in the conversion of lignocellulose biomass into ethanol for use in biofuels. Growth in media containing xylose as the sole carbon source enriches for genes encoding enzymes that better utilize this energy source. Using this strategy, monosaccharide transporters¹⁰⁶ and a xylose isomerase¹⁰⁷ were evolved for more efficient xylose consumption and ethanol production in *S. cerevisiae*.

The design of selections for protein activities that do not fulfil metabolic functions is more challenging and requires ingenuity. For example, Hilvert and colleagues¹⁰⁸ evolved nanocontainers to more effectively trap HIV protease, a protein that is toxic to *E. coli* hosts and for which sequestration confers faster growth rates. This approach yielded lumazine synthase capsids that had tenfold higher loading capacity.

Selections within in vitro compartments. *In vitro* selections can bypass limitations of *in vivo* selections such as transformation efficiency bottlenecks and host genome mutations that unexpectedly influence selection survival. A popular approach to couple genes and gene products without using cells is the translation of library members in artificial compartments such as the aqueous droplets of water–oil emulsions. Selections within *in vitro* compartments are particularly well suited for enzymes that directly act on DNA substrates. For example, in a selection for meganucleases with altered sequence specificity, Stoddard and colleagues¹⁰⁹ placed a mutated substrate sequence directly upstream of the meganuclease gene; DNA cleavage generated sticky ends that were competent for ligation of a PCR adapter. As a result, PCR within the emulsion droplets selectively amplified genes encoding nucleases that were active on the new substrate sequences.

Auxotroph complementation
The ability of functional library members to resolve a metabolic defect in the host, leading to replication of DNA that encodes active library members.

In vitro selections for DNA and RNA polymerases in emulsions are also referred to as compartmentalized self-replication (CSR) because the polymerases that most efficiently replicate their encoding gene in an emulsion PCR are enriched post-selection¹¹⁰ (FIG. 4b). Using CSR, Holliger and colleagues¹¹⁰ evolved DNA polymerases with higher thermostability and expanded substrate preferences, including Taq polymerase variants that accept Cy3 and Cy5 fluorophore-linked dNTPs¹¹¹. These evolved polymerases directly incorporate bright fluorescent dyes into DNA molecules, generating nucleic acid polymers with highly altered physical and chemical properties¹¹¹. In a separate study, Holliger and colleagues¹¹² used CSR to select DNA polymerases that more efficiently amplify damaged DNA isolated from extinct organisms.

The development of compartmentalized partnered replication (CPR) extends IVC selections beyond enzymes that act on DNA¹¹³. In CPR selection schemes,

the evolving enzymatic activity controls expression of Taq polymerase. Higher concentrations of Taq lead to better PCR amplification of active genes within emulsion droplets containing single *E. coli* cells (FIG. 4c). The first demonstrations of CPR evolved T7 RNA polymerase variants with orthogonal promoter preferences^{113,114}, an achievement that could in principle be accomplished using CSR. However, the power of CPR to evolve enzymes that do not act on DNA substrates was demonstrated through the evolution of tryptophanyl-tRNA synthetases that selectively charged the non-canonical amino acid 5-hydroxy-L-tryptophan onto suppressor tRNAs that suppress stop codons placed in the Taq polymerase gene¹¹³.

Emerging evolution paradigms

Continuous evolution. Traditional protein evolution methods require discrete time- and labour-intensive steps in which researchers generate gene libraries,

Table 2 | Comparison between screening and selection strategies

Method	Genotype–phenotype linkage	Reporter, mechanism of enrichment and detection method*	Library size	Evolvable phenotypes	Advantages	Disadvantages
Screening strategies						
Colonies on solid media	Spatial separation of variants	Visible signal from fluorescent or colorimetric gene product, or surrogate substrate; detected by manual inspection	Throughput limit: 10 ² –10 ⁴	Fluorescent proteins, binding interactions and a limited number of catalytic enzymes	Straightforward implementation	Low throughput and laborious; qualitative or semi-quantitative measurements
Isolated liquid cultures	Spatial separation of variants	Biochemical assays, visible signal from fluorescent gene product or surrogate substrate; detected by HPLC, NMR, MS or microplate reader	Throughput limit: 10 ² –10 ⁴	Fluorescent proteins and catalytic enzymes	Straightforward implementation; flexibility in choice of reporter and detection method; quantitative measurements; many classes of enzymes and proteins can be evolved	Low throughput and laborious
Cell surface display	Compartmentalized (cell membrane)	Non-diffusing fluorescent reporters (fluorescent-labelled antibodies or tags); detected by FACS	Throughput limit: 10 ⁸ (REFS 70,71,77)	Binding interactions, bond formation, bond cleavage and protein stability	High throughput; yeast display offers eukaryotic gene expression with chaperones and post-translational modifications	Limited applications to catalytic enzyme evolution (for example, transpeptidases ⁷⁷ , proteases ⁸⁵ , esterases ¹³⁶ and peroxidases ¹³⁷)
IVC	Compartmentalized (water–oil emulsions or polyelectrolyte shells)	Fluorescent gene product or fluorogenic substrate; detected by FACS	Throughput limit: 10 ⁷ –10 ⁸ (REFS 70,87,88)	Fluorescent proteins, catalytic enzymes for which a fluorogenic substrate can be made	High-throughput quantitative measurements	IVC and microfluidic techniques require expertise; components for emulsion and polyelectrolyte shells must be optimized for compatibility with a given fluorogenic reporter

introduce them into translation systems such as cells or *in vitro* compartments, perform screens or selections, and then isolate genes encoding library members with desired activities. Recently, researchers have developed methods by which all steps of the protein evolution cycle are performed continuously without manual intervention. These continuous evolution systems can markedly increase the efficiency of protein evolution and, therefore, the number of steps in the sequence space that can be explored in the search for optimal protein variants¹¹⁵.

The majority of continuous evolution experiments have selected for replicative fitness of microorganisms under continuous dilution. This continuous culture format has been applied to the evolution of bacterial genomes for shortened replication time¹¹⁶ and resistance to antibiotics¹¹⁷. Single-gene evolutions are also feasible in continuous culture, as demonstrated with chorismate mutase¹¹⁸ and β -lactamases¹¹⁹. However, specially designed continuous mutagenesis methods that only target the evolving gene of interest are crucial

Table 2 (cont.) | Comparison between screening and selection strategies

Method	Genotype–phenotype linkage	Reporter, mechanism of enrichment and detection method*	Library size	Evolvable phenotypes	Advantages	Disadvantages
Selection schemes						
Phage, mRNA, ribosome and cell surface display	Covalent	Binding to immobilized targets	Transformation limit: 10^{10} (phage display) ⁹³ ; throughput limit: 10^{12} – 10^{13} (mRNA and ribosome display) ^{97,98}	Binding interactions and a limited number of selections for catalysis ^{99,100,138}	Protocols for binding selection and affinity maturation are well established	Bacterial and <i>in vitro</i> expression biases; many eukaryotic proteins will not fold properly; not useful for evolving most enzymatic activities
Organismal fitness and auxotroph complementation	Compartmentalized (cell membrane)	Host fitness depends on evolving gene	Transformation limit: 10^8 (<i>S. cerevisiae</i>) ⁹⁵ ; 10^{10} (<i>E. coli</i>) ⁹⁶	Metabolic genes, selectable markers (antibiotic resistance) and suppressors of toxic genes	Conceptually simple	Strain optimizations can confound isolation of beneficial mutations in the evolving gene
IVC	Compartmentalized (water–oil emulsions)	Water–oil emulsions compartmentalize DNA, and <i>in vitro</i> transcription and translation yield proteins; selections for enzymes that act on their encoding DNA are common	Emulsion efficiency: 10^{10} (REF. 86)	Nucleases, DNase inhibitors, DNA ligases and binding interactions	Large library sizes; lacks transformation bottleneck	Heavily geared towards evolution of enzymes with DNA substrates; IVC requires technical expertise
CSR	Compartmentalized (water–oil emulsions)	DNA or RNA polymerases must amplify their encoding gene in an emulsion PCR	Emulsion efficiency: 10^{10} (REF. 86)	DNA and RNA polymerases	Large library sizes; lacks transformation bottleneck	By definition, this technique only applies to polymerases; IVC requires technical expertise
CPR	Compartmentalized (water–oil emulsions)	Emulsion PCR best amplifies genes that triggered the most Taq polymerase	Transformation limit and emulsion efficiency: 10^6 – 10^8 (<i>E. coli</i>) ^{96,113}	RNA polymerase and tRNA synthetase (hypothetically, anything that can be linked to gene expression)	Generalizable to many proteins	Challenging to design genetic circuits that link enzymatic activity to Taq expression; IVC requires technical expertise
PACE	Compartmentalized (cell membrane)	Expression of essential phage gene (such as gene III) is triggered by the desired phenotype; thus, infectious progeny are made in proportion to desired activity	Phage titre limit: 10^8 – 10^{12} (REF. 122)	RNA polymerase and protease (hypothetically, anything that can be linked to gene expression)	Generalizable to many proteins; continuous format enables ~100-fold more rounds of protein evolution per unit time than traditional methods	Challenging to design genetic circuits that link enzymatic activity to gene III expression; PACE requires technical expertise

*Reporter for screens or mechanism of enrichment for selections. CPR, compartmentalized partnered replication; CSR, compartmentalized self-replication; *E. coli*, *Escherichia coli*; FACS, fluorescence-activated cell sorting; HPLC, high-performance liquid chromatography; IVC, *in vitro* compartmentalization; MS, mass spectrometry; NMR, nuclear magnetic resonance; PACE, phage-assisted continuous directed evolution; *S. cerevisiae*, *Saccharomyces cerevisiae*.

for long evolutionary trajectories to avoid host genome mutations that circumvent selections by inducing cell survival for reasons unrelated to the protein of interest. For this reason, error-prone polymerases that exclusively replicate the library member are particularly amenable to continuous evolution in both *E. coli*¹¹⁹ and

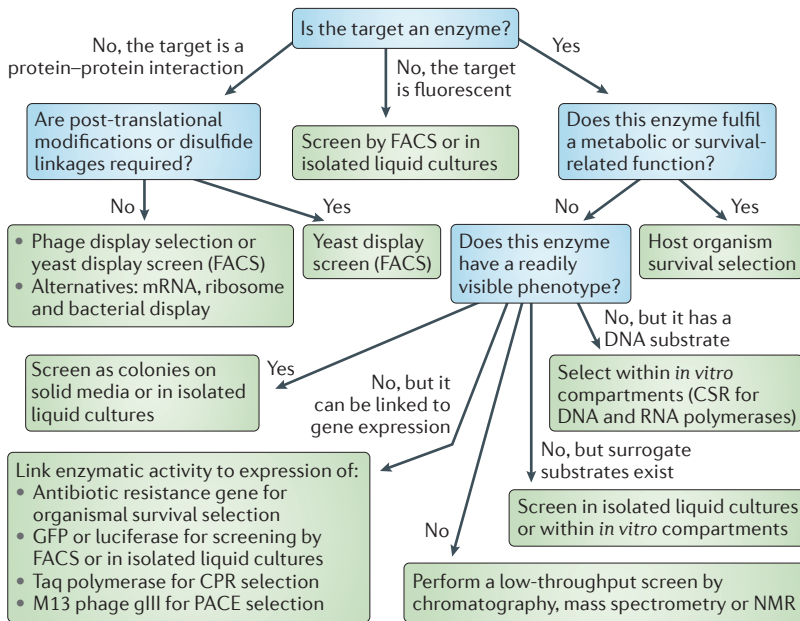
*S. cerevisiae*¹⁸. In addition, the aforementioned system for *in vivo* recombination in *S. cerevisiae* exclusively triggers recombination in an evolving gene during alternating stages of sporulation and selection⁴⁶.

Continuous evolution of viruses, including bacteriophage, is conducted in a fixed-volume vessel (a cellstat or 'lagoon') that is diluted with fresh bacterial host cells. The average residence time in the vessel is shorter than the time required for bacterial replication but longer than phage replication; thus, mutations only accumulate in the phage genome. This process has been used to study evolutionary dynamics within viral genomes^{120,121}, but our laboratory^{122–124} (also B.P. Hubbard and D.R.L., unpublished observations) has more recently extended its application to single-gene evolution. In our phage-assisted continuous evolution (PACE) system, an evolving gene is inserted into the M13 bacteriophage genome in place of an essential phage gene such as gene III (gIII). Instead, the evolving gene controls expression of gIII from an accessory plasmid. If the phage encodes a functional library member, then pIII, the protein encoded by gIII, is produced. Only phage assembled in the presence of pIII are infectious and can go on to infect and replicate in fresh host cells that dilute the vessel (FIG. 4d).

The continuous nature of PACE coupled with enhanced *in vivo* mutagenesis enables several hundred 'rounds' of selection, mutation and replication to take place per week without manual intervention. The first demonstration of PACE not only reprogrammed the promoter preferences of the T7 RNA polymerase but also suggested schemes for protein–protein interactions and recombinases¹²². A subsequent study developed a dominant-negative phage protein pIII-neg that can poison progeny phage and form the basis of negative selection¹²³. The recent use of PACE to continuously evolve proteases¹²⁴ and DNA-binding proteins (B.P. Hubbard and D.R.L., unpublished observations) demonstrates how PACE can be generalized through the development of gene circuitry that links desired enzymatic activities to the expression of gIII.

Computational design and directed evolution. Continuous evolution can extensively explore a fitness landscape over many rounds of evolution but, similar to other methods described above, accesses mutants that successively emerge from a starting gene. Computational protein design can initiate sequence space exploration from starting points that are inaccessible to evolutionary processes originating from naturally existing genes; as a result, it has the potential to expedite the evolution of completely novel protein functions^{125,126}. Although growing computational power and more sophisticated design methodologies have recently produced complex designs such as macromolecular assemblies, receptors and even catalysts^{127–129}, initial designs frequently remain suboptimal and require directed evolution to achieve high activity. For example, we and our collaborators¹³⁰ used phage and yeast display to increase affinity between the designed binding partners Pdar and Prb. Designed enzymes such as peroxidases¹³¹ and retroaldolases⁶¹ can also be optimized through evolution, yielding efficiencies

a Choosing a screening or selection method



b Choosing a library diversification strategy

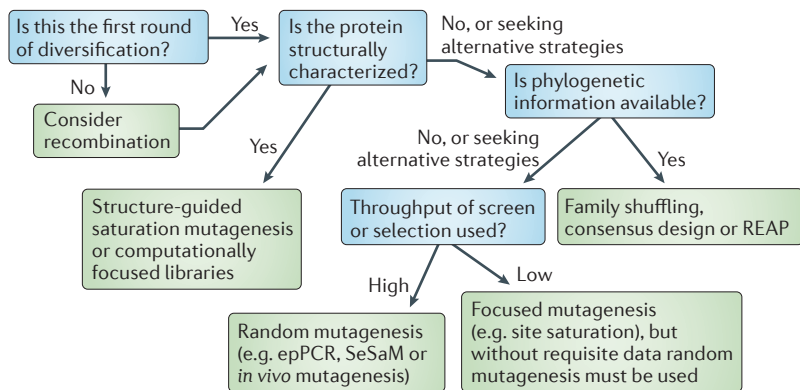


Figure 5 | Optimal strategies for directed evolution. **a** | The choice of a screening or selection method can be depicted as a decision tree that operates primarily on the properties of the protein and phenotype to be evolved. Although many techniques can be extended to alternative phenotypes, this figure focuses on the most popular methods for each set of conditions. **b** | Diversification strategies must be chosen both at the outset of an evolution project and between rounds of screening or selection. Considerations can and should change over the course of a project due to the phenotypes and genotypes within the evolving population. This decision tree attempts to distil these considerations with an emphasis on focused mutagenesis methods that have the maximum potential to identify functional variants. epPCR, error-prone PCR; CPR, compartmentalized partnered replication; CSR, compartmentalized self-replication; FACS, fluorescence-activated cell sorting; gIII, gene III; NMR, nuclear magnetic resonance; PACE, phage-assisted continuous evolution; REAP, reconstructed evolutionary adaptive path; SeSaM, sequence saturation mutagenesis.

that rival unrelated natural catalysts of the same reactions. Perhaps the most impressive testimony to the power of computational design coupled with directed evolution is the creation of novel protein catalysts. Tawfik, Baker and co-workers^{132,133} achieved this aim by designing and evolving proteins that catalyse the Kemp elimination, a reaction not known to be carried out by natural enzymes.

Conclusions and perspectives

Current protein evolution methods each offer unique features that make them more appropriate for solving certain classes of molecular problems (TABLE 2). When choosing a methodology, researchers should assess the features of the protein that is being evolved to find an optimal screening and selection technology, as well as an appropriate accompanying genetic diversification strategy (FIG. 5). Pioneering studies in the field of directed evolution sought to improve the wild-type activity of enzymes through the enhancement of solubility, thermostability, affinity for substrate or catalytic

turnover. These properties remain important in contemporary directed evolution because increased activity and stability often facilitate the engineering or evolution of other desirable properties. The pursuit of ambitious goals such as reprogrammed substrate selectivity^{33,85} and synthetically useful biocatalysts¹³⁴ benefits from innovative screens and selections that balance the need for throughput and accurate assessments of library members. New screens and selections that achieve higher throughput or carry out more continuous rounds of evolution can broaden the exploration of the fitness landscape, whereas novel mutagenesis strategies increase the search efficiency. Through computational techniques and creative molecular biology protocols, diversity is focused on residues and specific mutations that influence desired activities¹³⁵. New directed evolution methods will continue to generate proteins with useful new activities and specificities, as well as expand the scope of protein evolution to include even larger sets of chemical and biological functions.

- Wright, S. I. *et al.* The effects of artificial selection on the maize genome. *Science* **308**, 1310–1314 (2005).
- Driscoll, C. A., Macdonald, D. W. & O'Brien, S. J. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl Acad. Sci. USA* **106** (Suppl. 1), 9971–9978 (2009).
- Umeno, D., Tobias, A. V. & Arnold, F. H. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol. Mol. Biol. Rev.* **69**, 51–78 (2005).
- Atsumi, S. & Liao, J. C. Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl. Environ. Microbiol.* **74**, 7802–7808 (2008).
- Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
- Zhang, Y. X. *et al.* Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**, 644–646 (2002).
- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, C. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).
- Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
- Mclsaac, R. S. *et al.* Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl Acad. Sci. USA* **111**, 13034–13039 (2014).
- Jespersen, L. S., Roberts, A., Mahler, S. M., Winter, G. & Hoogenboom, H. R. Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Biotechnology* **12**, 899–903 (1994).
- Lai, Y. P., Huang, J., Wang, L. F., Li, J. & Wu, Z. R. A new approach to random mutagenesis *in vitro*. *Biotechnol. Bioeng.* **86**, 622–627 (2004).
- Myers, R. M., Lerman, L. S. & Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. *Science* **229**, 242–247 (1985).
- Freese, E. Specific mutagenic effect of base analogues on Phage-T4. *J. Mol. Biol.* **1**, 87–105 (1959).
- Bridges, B. A. & Woodgate, R. Mutagenic repair in *Escherichia coli*: products of the *recA* gene and of the *umuD* and *umuC* genes act at different steps in UV-induced mutagenesis. *Proc. Natl Acad. Sci. USA* **82**, 4193–4197 (1985).
- Cox, E. C. Bacterial mutator genes and the control of spontaneous mutation. *Annu. Rev. Genet.* **10**, 135–156 (1976).
- Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an *E. coli* mutator strain. *Mol. Biotechnol.* **7**, 189–195 (1997).
- Scheuermann, R., Tam, S., Burgers, P. M. J., Lu, C. & Echols, H. Identification of the ϵ -subunit of *Escherichia coli* DNA polymerase III holoenzyme as the *dnaQ* gene product: a fidelity subunit for DNA replication. *Proc. Natl Acad. Sci. USA* **80**, 7085–7089 (1983).
- Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
- Leung, D. W., Chen, E. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11–15 (1989).
- Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255**, 589–603 (1996).
- Eckert, K. A. & Kunkel, T. A. High fidelity DNA synthesis by the *Thermus Aquaticus* DNA polymerase. *Nucleic Acids Res.* **18**, 3739–3744 (1990).
- Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939–942 (2008).
- Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
This seminal study in optimizing the conditions for epPCR is a must-read for all scientists performing random mutagenesis.
- Vanhercke, T., Ampe, C., Tirry, L. & Denolf, P. Reducing mutational bias in random protein libraries. *Anal. Biochem.* **339**, 9–14 (2005).
- Wong, T. S., Tee, K. L., Hauer, B. & Schwaneberg, U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* **32**, e26 (2004).
- Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**, 315–323 (1985).
- Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–341 (2009).
- Quan, J. Y. & Tian, J. D. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 (2009).
- Nour-Eldin, H. H., Geu-Flores, F. & Halkier, B. A. User cloning and user fusion: the ideal cloning techniques for small and big laboratories. *Methods Mol. Biol.* **643**, 185–200 (2010).
- Reidhaaroolson, J. F. & Sauer, R. T. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* **241**, 53–57 (1988).
- Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* **1543**, 408–415 (2000).
- Chen, F. *et al.* Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl Acad. Sci. USA* **107**, 1948–1953 (2010).
- Cherny, I. *et al.* Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. *ACS Chem. Biol.* **8**, 2394–2403 (2013).
This paper nicely demonstrates how computational modelling can identify beneficial mutations, which can be stochastically incorporated into gene libraries.
- Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
- Wijma, H. J. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58 (2014).
- Herman, A. & Tawfik, D. S. Incorporating synthetic oligonucleotides via gene reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226 (2007).
- Stemmer, W. P. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
This study is the first to establish a method for homologous recombination of evolving protein populations.
- Coco, W. M. *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19**, 354–359 (2001).
- Muller, K. M. *et al.* Nucleotide exchange and excision technology (NEXT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Res.* **33**, e117 (2005).
- Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).
- Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
- Ness, J. E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251–1255 (2002).
- Zha, D. X., Eipper, A. & Rietz, M. T. Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. *ChemBiochem* **4**, 34–39 (2003).
- Cramer, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315–319 (1996).
- Cramer, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).

46. Romanini, D. W., Peralta-Yahya, P., Mondol, V. & Cornish, V. W. A. Heritable recombination system for synthetic Darwinian evolution in yeast. *ACS Synth. Biol.* **1**, 602–609 (2012).
47. Sieber, V., Martinez, C. A. & Arnold, F. H. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**, 456–460 (2001).
48. Ostermeier, M., Shim, J. H. & Benkovic, S. J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17**, 1205–1209 (1999).
49. Bittker, J. A., Le, B. V., Liu, J. M. & Liu, D. R. Directed evolution of protein enzymes using nonhomologous random recombination. *Proc. Natl Acad. Sci. USA* **101**, 7011–7016 (2004).
50. Voigt, C. A., Martinez, C., Wang, Z. C., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9**, 553–558 (2002).
51. Hiraga, K. & Arnold, F. H. General method for sequence-independent site-directed chimeragenesis. *J. Mol. Biol.* **330**, 287–296 (2003).
52. Kolkman, J. A. & Stemmer, W. P. C. Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* **19**, 423–428 (2001).
53. 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).
54. Gillam, E. M. J. *Directed Evolution Library Creation* (Springer, 2014).
This book is an excellent resource for comparing and choosing between genetic diversification methods as well as for successfully executing library generation protocols.
55. You, L. & Arnold, F. H. Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**, 77–83 (1996).
56. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* **91**, 12501–12504 (1994).
57. Kralj, J. M., Douglass, A. D., Hochbaum, D. R., Maclaurin, D. & Cohen, A. E. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nat. Methods* **9**, 90–130 (2012).
58. Cali, J. J. *et al.* Luminogenic cytochrome P450 assays. *Expert Opin. Drug Metab. Toxicol.* **2**, 629–645 (2006).
59. Ostafe, R., Prodanovic, R., Lloyd Ung, W., Weitz, D. A. & Fischer, R. A high-throughput cellulase screening system based on droplet microfluidics. *Biomicrofluidics* **8**, 041102 (2014).
60. Gupta, R. D. *et al.* Directed evolution of hydrolases for prevention of G-type nerve agent intoxication. *Nat. Chem. Biol.* **7**, 120–125 (2011).
61. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
62. Goddard, J. P. & Raymond, J. L. Enzyme assays for high-throughput screening. *Curr. Opin. Biotechnol.* **15**, 314–322 (2004).
63. Fields, S. & Song, O. K. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
64. Baker, K. *et al.* Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc. Natl Acad. Sci. USA* **99**, 16537–16542 (2002).
65. Lin, H. N., Tao, H. Y. & Cornish, V. W. Directed evolution of a glycosynthase via chemical complementation. *J. Am. Chem. Soc.* **126**, 15051–15059 (2004).
66. Peralta-Yahya, P., Carter, B. T., Lin, H. N., Tao, H. Y. & Cornish, V. W. High-throughput selection for cellulase catalysts using chemical complementation. *J. Am. Chem. Soc.* **130**, 17446–17452 (2008).
67. Swe, P. M. *et al.* Targeted mutagenesis of the *Vibrio fischeri* flavin reductase FRase I to improve activation of the anticancer prodrug CB1954. *Biochem. Pharmacol.* **84**, 775–783 (2012).
68. Sengupta, D., Lin, H. N., Goldberg, S. D., Mahal, J. J. & Cornish, V. W. Correlation between catalytic efficiency and the transcription read-out in chemical complementation: a general assay for enzyme catalysis. *Biochemistry* **43**, 3570–3581 (2004).
69. Fulwyler, M. J. Electronic separation of biological cells by volume. *Science* **150**, 910–911 (1965).
70. Shapiro, H. M. *Practical Flow Cytometry* (Wiley-Liss, 2005).
71. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
This paper describes the invention of yeast display protein libraries for screening protein–protein interactions and serves as the foundation for many other cell surface display methods.
72. Santoro, S. W. & Schultz, P. G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl Acad. Sci. USA* **99**, 4185–4190 (2002).
73. Wang, J. D., Herman, C., Tipton, K. A., Gross, C. A. & Weissman, J. S. Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell* **111**, 1027–1039 (2002).
74. Peck, S. H., Chen, I. & Liu, D. R. Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chem. Biol.* **18**, 619–630 (2011).
75. Rajpal, A. *et al.* A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc. Natl Acad. Sci. USA* **102**, 8466–8471 (2005).
76. Wang, X. X., Cho, Y. K. & Shusta, E. V. Mining a yeast library for brain endothelial cell-binding antibodies. *Nat. Methods* **4**, 143–145 (2007).
77. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl Acad. Sci. USA* **108**, 11399–11404 (2011).
78. Qu, Z. *et al.* Immobilization of actively thromboresistant assemblies on sterile blood-contacting surfaces. *Adv. Healthc. Mater.* **3**, 30–35 (2014).
79. Shi, J. H. *et al.* Engineered red blood cells as carriers for systemic delivery of a wide array of functional probes. *Proc. Natl Acad. Sci. USA* **111**, 10131–10136 (2014).
80. Ling, J. J., Policarpo, R. L., Rabideau, A. E., Liao, X. & Pentelute, B. L. Protein thioester synthesis enabled by sortase. *J. Am. Chem. Soc.* **134**, 10749–10752 (2012).
81. McCluskey, A. J. & Collier, R. J. Receptor-directed chimeric toxins created by sortase-mediated protein fusion. *Mol. Cancer Ther.* **12**, 2273–2281 (2013).
82. Policarpo, R. L. *et al.* Flow-based enzymatic ligation by sortase A. *Angew. Chem. Int. Ed. Engl.* **53**, 9205–9208 (2014).
83. Swee, L. K., Lourido, S., Bell, G. W., Ingram, J. R. & Ploegh, H. L. One-step enzymatic modification of the cell surface redirects cellular cytotoxicity and parasite tropism. *ACS Chem. Biol.* (2014).
84. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc. Natl Acad. Sci. USA* **111**, 13343–13348 (2014).
85. Yi, L. *et al.* Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc. Natl Acad. Sci. USA* **110**, 7229–7234 (2013).
86. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652–656 (1998).
The authors of this paper developed IVC as a platform for directed evolution. This study describes a selection for methyltransferases within water–oil emulsion droplets.
87. Bernath, K. *et al.* *In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal. Biochem.* **325**, 151–157 (2004).
88. Agresti, J. J. *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl Acad. Sci. USA* **107**, 4004–4009 (2010).
89. Scott, D. J. & Plückthun, A. Direct molecular evolution of detergent-stable G protein-coupled receptors using polymer encapsulated cells. *J. Mol. Biol.* **425**, 662–677 (2013).
90. Fischelchner, M. *et al.* Evolution of enzyme catalysts caged in biomimetic gel-shell beads. *Nat. Chem.* **6**, 791–796 (2014).
In this study, polyelectrolyte shells served as *in vitro* compartments for screening by flow cytometry.
91. Bessette, P. H., Rice, J. J. & Daugherty, P. S. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* **17**, 731–739 (2004).
92. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).
In this pioneering study, phage display is demonstrated as a powerful technique to select high-affinity antibody fragments. This paper also nicely illustrates the guiding principles of related binding enrichments.
93. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. Making antibody fragments using phage display libraries. *Nature* **352**, 624–628 (1991).
94. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
95. Becker, D. M. & Guarente, L. High-efficiency transformation of yeast by electroporation. *Methods Enzymol.* **194**, 182–187 (1991).
96. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
97. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA* **94**, 4937–4942 (1997).
98. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl Acad. Sci. USA* **98**, 3750–3755 (2001).
99. Amstutz, P. *et al.* *In vitro* selection for catalytic activity with ribosome display. *J. Am. Chem. Soc.* **124**, 9396–9403 (2002).
100. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828–831 (2007).
101. Orenica, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* **8**, 238–242 (2001).
102. Palmer, A. C. & Kishony, R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* **14**, 243–248 (2013).
103. Liu, D. R., Magliery, T. J., Pasternak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc. Natl Acad. Sci. USA* **94**, 10092–10097 (1997).
This groundbreaking study on genetic code expansion exemplifies how selectable antibiotic resistance markers can form the basis for a range of *in vivo* selections.
104. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044–1048 (2002).
105. Gaj, T., Mercer, A. C., Gersbach, C. A., Gordley, R. M. & Barbas, C. F. 3rd Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl Acad. Sci. USA* **108**, 498–503 (2011).
106. Young, E. M., Tong, A., Bui, H., Spofford, C. & Alper, H. S. Rewiring yeast sugar transporter preference through modifying a conserved protein motif. *Proc. Natl Acad. Sci. USA* **111**, 131–136 (2014).
This study uses an auxotroph complementation strategy to select for sugar transporters that selectively uptake xylose from culture media.
107. Lee, S. M., Jellison, T. & Alper, H. S. Directed evolution of xylose isomerase for improved xylose catabolism and fermentation in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **78**, 5708–5716 (2012).
108. Worsdorfer, B., Woycechowsky, K. J. & Hilvert, D. Directed evolution of a protein container. *Science* **331**, 589–592 (2011).
109. Takeuchi, R., Choi, M. & Stoddard, B. L. Redesign of extensive protein–DNA interfaces of meganucleases using iterative cycles of *in vitro* compartmentalization. *Proc. Natl Acad. Sci. USA* **111**, 4061–4066 (2014).
110. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA* **98**, 4552–4557 (2001).
111. Ramsay, N. *et al.* CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* **132**, 5096–5104 (2010).
112. d’Abadie, M. *et al.* Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25**, 939–943 (2007).

113. Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32**, 97–101 (2014).
The authors of this paper evolved enzymes within IVCs by linking the desired phenotype to the expression of Taq polymerase within *E. coli*. Taq can then be used in PCR to amplify the DNA encoding active library members within the emulsion droplet.
114. Meyer, A. J., Ellefson, J. W. & Ellington, A. D. Directed evolution of a panel of orthogonal T7 RNA polymerase variants for *in vivo* or *in vitro* synthetic circuitry. *ACS Synth. Biol.* <http://dx.doi.org/10.1021/sb500299c> (2014).
115. Badran, A. H. & Liu, D. R. *In vivo* continuous directed evolution. *Curr. Opin. Chem. Biol.* **24**, 1–10 (2015).
116. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).
117. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105 (2012).
118. Muller, M. M. *et al.* Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. *PLoS Genet.* **9**, e1003187 (2013).
119. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl Acad. Sci. USA* **100**, 9727–9732 (2003).
120. Bull, J. J. *et al.* Exceptional convergent evolution in a virus. *Genetics* **147**, 1497–1507 (1997).
121. Wichman, H. A., Wichman, J. & Bull, J. J. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* **170**, 19–31 (2005).
122. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
This study establishes a technological platform for the continuous evolution of biomolecules by linking the phage life cycle to the desired enzymatic activity.
123. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014).
124. Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* **5**, 5352 (2014).
125. Jiang, L. *et al.* *De novo* computational design of retroaldol enzymes. *Science* **319**, 1387–1391 (2008).
126. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science* **329**, 309–313 (2010).
127. Joh, N. H. *et al.* *De novo* design of a transmembrane Zn²⁺-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
128. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
129. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **346**, 1525–1528 (2014).
130. Karanicolas, J. *et al.* A *de novo* protein binding pair by computational design and directed evolution. *Mol. Cell* **42**, 250–260 (2011).
131. Patel, S. C. & Hecht, M. H. Directed evolution of the peroxidase activity of a *de novo*-designed protein. *Protein Eng. Des. Sel.* **25**, 445–452 (2012).
132. Khersonsky, O. *et al.* Optimization of the *in-silico*-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412 (2011).
133. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
This paper describes the computational design of a Kemp elimination catalyst. Subsequent screening yielded improved catalysts for a reaction that is not known to be performed by natural enzymes.
134. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
135. Lutz, S. & Patrick, W. M. Novel methods for directed evolution of enzymes: quality, not quantity. *Curr. Opin. Biotechnol.* **15**, 291–297 (2004).
136. Becker, S. *et al.* Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew. Chem. Int. Ed Engl.* **47**, 5085–5088 (2008).
137. Lipovsek, D. *et al.* Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display. *Chem. Biol.* **14**, 1176–1185 (2007).
138. Piotukh, K. *et al.* Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J. Am. Chem. Soc.* **133**, 17536–17539 (2011).

Acknowledgements

This work was supported by the US Defense Advanced Research Projects Agency grants DARPA HR0011-11-2-0003 and DARPA N66001-12-C-4207, the US National Institutes of Health (NIH)/National Institute of General Medical Sciences (NIGMS) (grant R01 GM095501) and the Howard Hughes Medical Institute (HHMI).

Competing interests statement

The authors declare no competing interests.