CHEMICAL REVIEWS

pubs.acs.org/CR Review

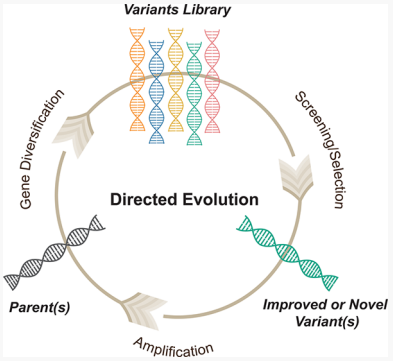
Directed Evolution: Methodologies and Applications

Yajie Wang,[#] Pu Xue,[#] Mingfeng Cao, Tianhao Yu, Stephan T. Lane, and Huimin Zhao*

Cite This: *Chem. Rev.* 2021, 121, 12384–12444 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Directed evolution aims to expedite the natural evolution process of biological molecules and systems in a test tube through iterative rounds of gene diversifications and library screening/selection. It has become one of the most powerful and widespread tools for engineering improved or novel functions in proteins, metabolic pathways, and even whole genomes. This review describes the commonly used gene diversification strategies, screening/selection methods, and recently developed continuous evolution strategies for directed evolution. Moreover, we highlight some representative applications of directed evolution in engineering nucleic acids, proteins, pathways, genetic circuits, viruses, and whole cells. Finally, we discuss the challenges and future perspectives in directed evolution.



CONTENTS

1. INTRODUCTION

Evolution is the hallmark of life, a key concept brought to light by Darwin's *On the Origin of Species* in 1859. Prior to that, humankind had harnessed the power of evolution to create target organisms with desired traits through selective breeding and domestication for centuries, but without the knowledge of the underlying evolutionary principles and processes. In the

Received: March 28, 2021
Published: July 23, 2021

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384–12444

late 1960s and early 1970s, Spiegelman and co-workers performed a series of *in vitro* RNA replication experiments under different selective pressures to explore the fundamental evolutionary principles.[1−3] In 1985, Smith developed the phage display method to enrich peptides with desired binding properties, which led to numerous applications in antibody engineering.[4] Around the same time, the concept of the "evolutionary machine" for evolving proteins by mutagenesis and selection was formally proposed by Eigen and Gardiner;[5] and the experiments performed by Szostak and Hageman set the stage for directed enzyme evolution.[6,7] However, this concept was not formally reduced to practice by Arnold and co-workers until 1993, which marked the beginning of the directed evolution field even though this concept has occasionally been used to describe adaptive evolutionary experiments for decades.

Directed evolution mimics Darwinian evolution in a test tube and comprises iterative cycles of generating genetic diversity followed by screening/selection. Unlike natural evolution, whose goal is survival and reproduction, directed evolution is executed at much higher mutation and recombination rates to screen for the desired biological function. The general procedure of directed evolution includes two main steps: (1) gene diversification by random mutagenesis and/or gene recombination to generate a diverse library of variants; and (2) screening/selection to obtain variants with improved phenotypes (Figure 1). The improved

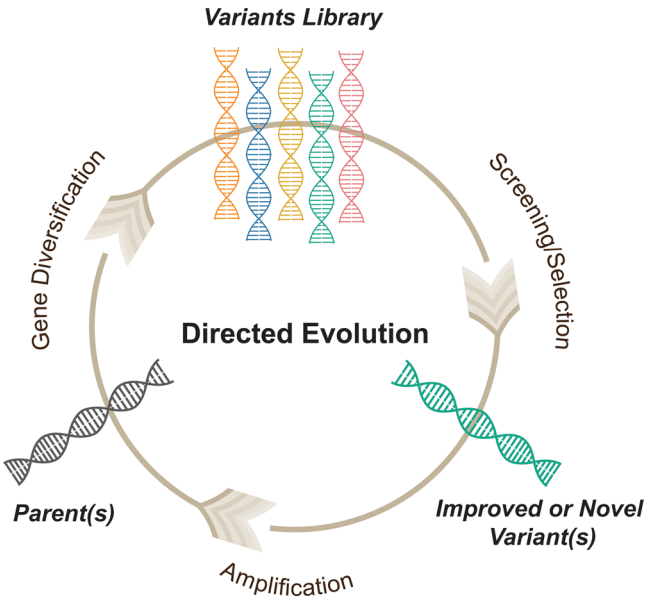


Figure 1. Principle of directed evolution.

variant will serve as a new starting point for the next round of gene diversification. Iterative rounds of directed evolution can be executed until the desired variants "climb to the top of the hill".

Like many other scientific fields, the establishment of directed evolution was enabled by technological advancements and exploration. Figure 2 summarizes the various important milestones in directed evolution. Early applications of directed evolution mainly focused on evolving individual proteins of interest for improved stability, activity, selectivity, and compatibility with industrial process conditions.[8−10] More recently, directed evolution was used to engineer promiscuous enzymes for abiotic transformations.[11−14] In addition to enzymes, directed evolution has been extended to engineer binding proteins and nucleic acids, and extended to evolve genetic circuits, biochemical pathways, and even whole genomes for biotechnological and biomedical applications.[15−17]

In this review, we describe the most frequently used methodologies for directed evolution, including gene diversification strategies, screening/selection methods, and recently developed strategies for continuous evolution. In addition, we highlight some representative applications of directed evolution in engineering nucleic acids, proteins, pathways, genetic circuits, viruses, and whole cells. Finally, we briefly discuss the challenges, opportunities, and future directions of directed evolution.

2. METHODOLOGIES

In this section, we provide a comprehensive review on the gene diversification, selection, and screening methods developed for directed evolution in the past three decades.

2.1. Methods for Gene Diversification

Genes can be diversified in either a completely random or a relatively targeted manner. Considering each amino acid can have 19 alternatives, preparing a comprehensive library to cover the entire mutational space of a protein is never practical. Random mutagenesis of a single gene can perform optimal sparse sampling of the sequence space to identify the "hot spots" that are largely correlated to the desired protein property without detailed structure or function information. Focused mutagenesis, on the other hand, maximizes the sampling of certain positions that determine the protein function. Variants with improved properties can be identified from a much smaller focused mutagenesis library only if the correlation between the sequence and function is clear. *In vitro* mutagenesis is the most common strategy for well controlled and efficient gene diversification. However, thanks to recent advances in gene editing tools, *in vivo* mutagenesis is becoming more popular with great potential to be performed continuously without human intervention. The size of the sequence space that can be sampled directly determines the success rate of a directed evolution experiment. To minimize the sequence space to be sampled and expedite the evolutionary process, computational modeling and machine learning can be used to predict beneficial mutations and can allow researchers to create smaller, more efficient variant libraries.

2.1.1. *In Vitro* Gene Diversification. *In vitro* polymerase chain reaction (PCR)-based techniques such as error-prone PCR (epPCR), site saturation mutagenesis (SSM), and recombination-based DNA shuffling are the most widely used tools for mutagenizing the parent genes (Figure 3). EpPCR mimics the imperfect DNA replication process occurring naturally (10^{-10} per base pair (bp)) by introducing random mutations with a much higher mutation rate (10^{-4} per bp) while copying the target genes. It enables rapid engineering of a target protein without requiring knowledge of the protein structure−function relationships. When the target protein's structure is available, focused mutagenesis can be performed on selected residues contributing to the enzyme's stability, activity, and selectivity, thus creating a smaller but more efficient library to ease selection and increase the likelihood of obtaining improved variants.[18,19] DNA

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

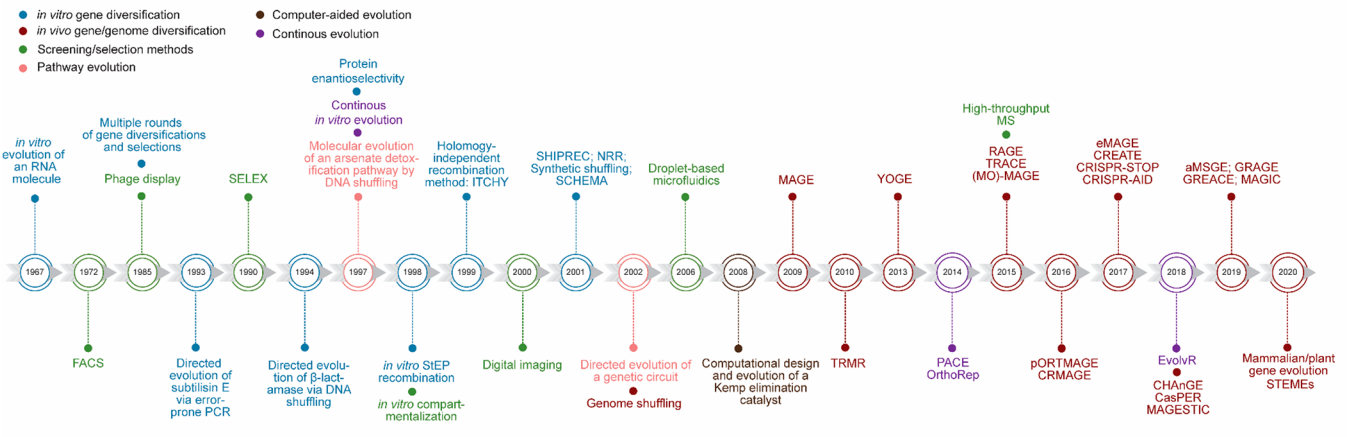


Figure 2. Key milestones of directed evolution.

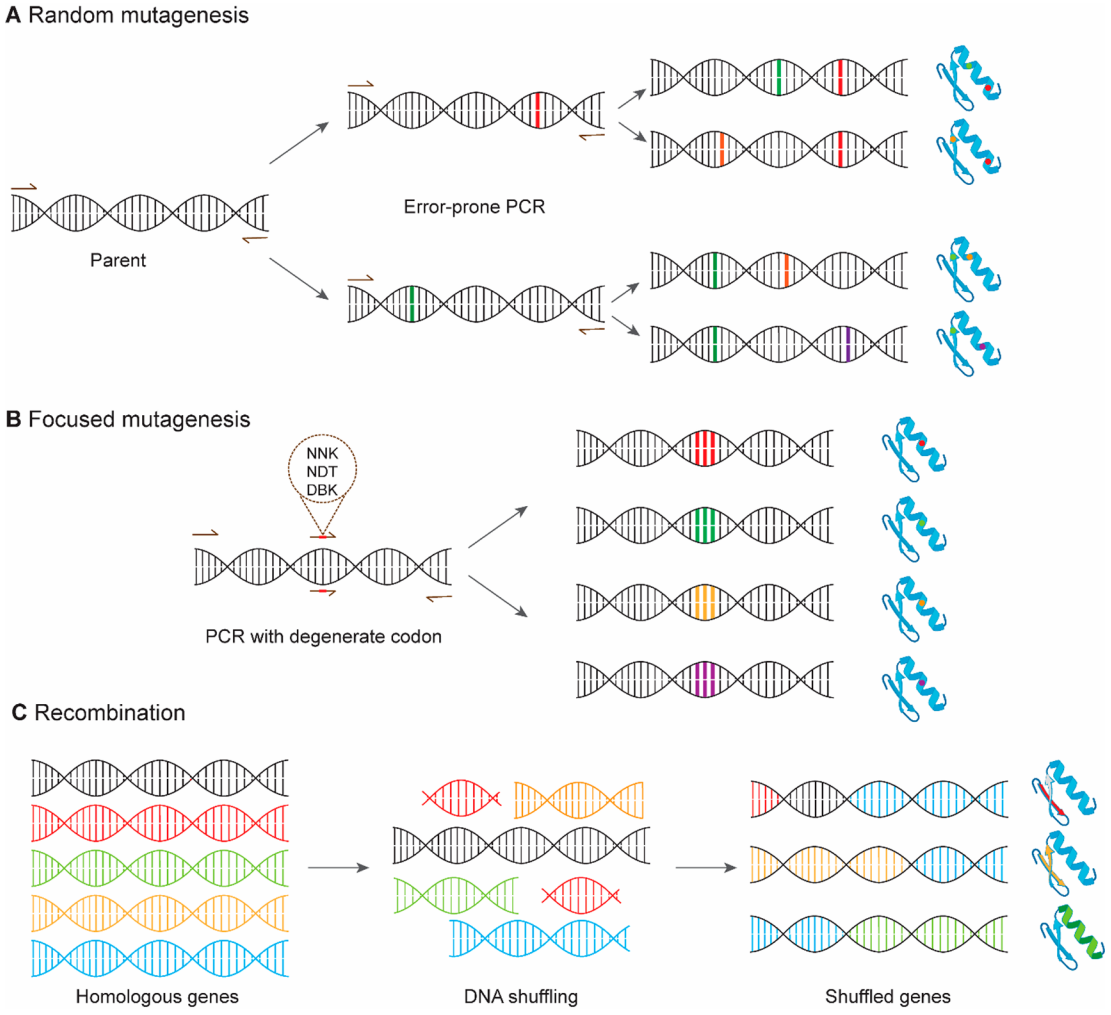


Figure 3. Summary of *in vivo* gene diversification methods. (A) Random mutagenesis to introduce point mutations by using the error-prone PCR method.[20] (B) Focused mutagenesis by using overlap extension PCR with degenerate oligonucleotides.[46] (C) Recombination by DNA shuffling.[68,69]

shuffling, on the other hand, imitates the natural process of homologous recombination to access combinations of beneficial mutations by mutagenizing and recombining homologous genes at crossover points with high sequence similarity. Compared with traditional random mutagenesis methods using chemical or physical agents to randomly damage DNA, *in vitro* gene diversification methods are more controlled processes that exhibit higher mutation rates and reduced mutational bias. Random mutagenesis, SSM and DNA recombination can be used separately or together to achieve different directed evolution goals (Table 1).

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 1. Summary of *In Vitro* Gene Diversification Methods

category	methods	principles	advantages	limitations	ref
Random and focused mutagenesis	EpPCR	Error-prone polymerase dependent random point mutation	Require no structure and function information; easy to use	Mutation bias and narrow chemistry diversity; require high-throughput screening/selection	20−25
	SeSaM	Polymerase independent random point mutation	Introduce consecutive mutations	Laborious; require synthesized nucleotides	26−29
	RID	Random insertion/deletion (up to 16 bps) on the whole plasmid	Introduce indel mutations	Laborious; whole plasmid mutations	30
	Transposon-aided	Mu-guided gene disruption followed by *Mly*I digestion and self-ligation	Introduce random trinucleotide deletion, insertion, and domain insertion on target gene	Lead truncated proteins; require high-throughput screening/selection	31
	SSM	Saturated mutagenesis on selected residues by PCR with synthetic DNA oligonucleotides containing one or more degenerate codons	Smaller but smarter library; broad chemistry diversity	Require detailed structure or function information	32−43, 45
Recombination-based mutagenesis	DNA Shuffling	HR-dependent DNA shuffling: DNAs are fragmented by DNase I or restriction enzymes, followed by primer-free PCR cycles	Access sequential space inaccessible by random and focused mutagenesis	Parent strands contamination; low diversity of the chimeric library	68, 69
	RACHITT	HR-dependent DNA shuffling: ssDNA is used to prepare fragment library and uracil-containing transit ssDNA template is used to assist assembly	No parent strands contamination; higher crossover frequency at less identified regions	Require uracil-containing ssDNA template	71
	StEP	HR-dependent DNA shuffling: special PCR cycles with elongation being disrupted earlier	Fragmentation step free; easy to use	Challenging on long genes with low sequence similarity	73
	Synthetic Shuffling	HR-dependent DNA shuffling: primer extension by using degenerate oligonucleotides covering the parent genes	Diverse chimeric sequence libraries	Challenging on long genes with low sequence similarity	67
	ITCHY	HR-independent DNA shuffling: gene is truncated by restriction enzyme, followed by blunt end ligation	Create fusion library between two genes with low similarity; introduce fusion points at nonhomology sites	Lead truncated proteins; limited to two-crossover libraries	76
	SHIPREC	Modified version of ITCHY: fuse two gene head-to-tail before random truncation	Reduced truncated genes in the final library	Limited to two-crossover libraries	78
	NRR	HR-independent DNA shuffling: gene is truncated by DNase I followed by blunt ligation	Generate diverse libraries with up to 11 crossovers	Lead nonfunctional proteins.	79

2.1.1.1. Random Mutagenesis. Mutation rate and the mutational spectrum are the most important factors to be considered when designing a random mutagenesis experiment. A typical random mutagenesis method has a moderate mutation rate of 10^{-4} to 10^{-3} substitutions per replicated base (s.p.b.). An unbiased method exhibits a transition (Ts) to transversion (Tv) ratio of 0.5, AT → GC to GC → AT transition mutation ratio of 1, and the equal frequency of mutating A's and T's versus the frequency of mutating G's and C's.

EpPCR was first developed by Goeddel and co-workers in 1989 and remains the most commonly used *in vitro* random mutagenesis method.[20] The basic principle is to use a low fidelity DNA polymerase such as *Taq* and Mutazyme polymerases to generate point mutations during PCR amplifications under nonstandard conditions. Increased magnesium concentration, supplementation with manganese, unbalanced dNTP concentrations, or extended PCR cycles can reduce the fidelity of base-paring and increase the mutation rates up to 8×10^{-3} s.p.b. Using mutagenic nucleotide analogues can further improve the mutation rates up to 10^{-2}∼10^{-1}s.p.b.[21] The traditional epPCR using *Taq* polymerase has a bias for AT → GC transitions and AT → TA transversions. A more balanced mutation distribution can be created by using a blend of *Taq* polymerase and Mutazyme DNA polymerase that preferentially generates GC → AT transitions and GC → TA transversion mutations.[22] This strategy is commercially adopted for preparing a random mutagenesis kit. To simplify the general cloning process involved in epPCR, Hayashi and co-workers developed error-prone rolling circle amplification to generate a randomly mutated plasmid library with a moderate mutation rate of 3×10^{-3} s.p.b.[23−25] The key advantage of this technique is the simple procedure without requiring either PCR or restriction digestion-ligation cloning experiments. However, it is less controlled since mutations can be generated along the whole plasmid.

Sequence saturation mutagenesis (SeSaM) is a polymerase independent random mutagenesis tool complementary to epPCR. It has been found that beneficial substitutions can be preferentially achieved by having a strong bias for transversions during mutagenesis and by introducing consecutive mutations.[26] The combination of transition bias and inability to create adjacent mutations in the epPCR largely limits the genetic diversity. The basic principle of SeSaM is to use α-phosphorothioate nucleotides to "select" any nucleotide to mutate and use universal bases that have promiscuous base pairing ability and coding bias to achieve mutations with tunable mutational bias.[26,27] Transversion-enriched SeSaM led to G → T (∼20%) and G → C (8%) transversion mutations with Ts/Tv ∼0.5 and consecutive mutations (up to 37%) that were hardly ever achieved by epPCR.[28,29]

Indel mutations, amino acid deletions and insertions, are inaccessible by epPCR and SeSaM, but are commonly observed among natural protein homologues and can result in conformational and functional variants that cannot be obtained through side-chain substitutions. Sisido and co-workers first reported a random insertion/deletion mutagenesis (RID) method that enabled deletion of an arbitrary number of consecutive bases (up to 16 bps) at a random position while inserting an arbitrary number of specific or random bases into the same position.[30] However, its experimental implementation was relatively tedious. Different

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

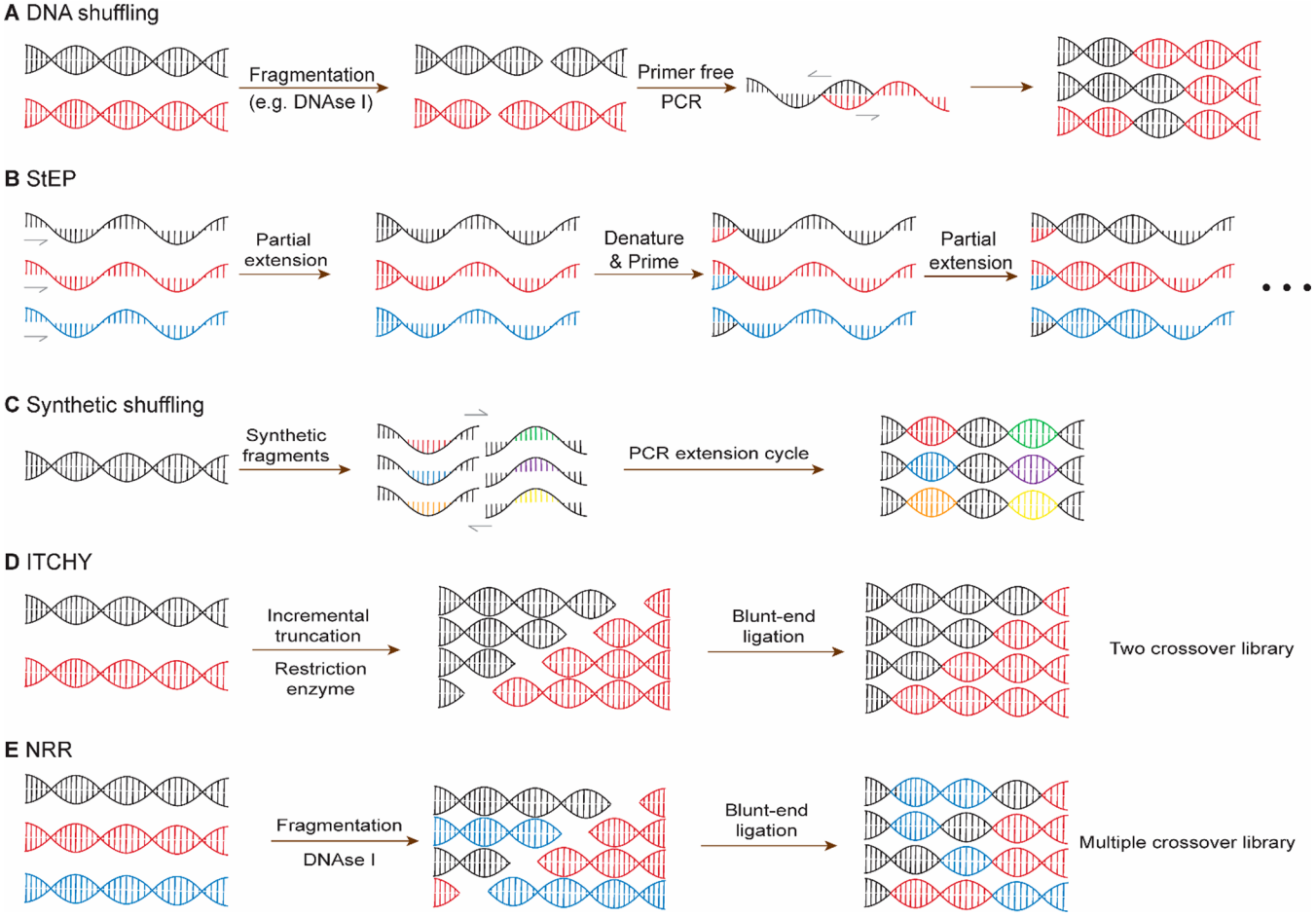


Figure 4. Representative combination-based gene diversification methods *in vitro*.[67−69,73,76,79]

from RID which introduced indel mutations randomly to the whole plasmid, transposon-based approaches were developed to sample trinucleotide deletions, trinucleotide replacements, and domain insertions on specific genes of interest.[31] All transposon-aided approaches shared the same principle to introduce indel mutations: gene disruption by transposition with an engineered Mu transposon, followed by *Mly*I digestion and self-ligation. The experiment is convenient to carry out, but the created library contains many truncated proteins; thus high-throughput screening methods are necessary for the successful application of this strategy.[667]

2.1.1.2. Focused Mutagenesis. SSM is the most popular focused mutagenesis method in which a single amino acid residue can be substituted with all other 19 amino acid residues. Selected residues can also be clustered together for mutagenesis to increase the dimension of library complexity. All SSM approaches introduce mutations during oligo annealing or PCR by using synthetic DNA oligonucleotides containing one or more degenerate codons at the target residues. SSM libraries can be generated by a couple of different methods: (1) cassette mutagenesis; (2) whole plasmid amplification; and (3) PCR mutagenesis.

In cassette mutagenesis, a small region around the target site is excised from the plasmid. The same mutagenized region flanked by corresponding restriction sites is prepared by synthetic oligonucleotide annealing and introduced in the plasmid by restriction digestion and ligation. This method generates a library of variants with low wild-type protein background since the protein function will only be restored when the mutagenized cassette is inserted. However, it is limited to prepare single-site mutagenesis libraries or multisite mutagenesis libraries with target residues close to each other, i.e., within 60 bp nucleotides.[32−34] Whole plasmid amplification is a very simple method without the need for any subcloning. A pair of complementary[35] or partial 5′ overlap primers[36] containing degenerate codons is used to amplify the whole plasmid harboring the gene of interest. The resultant library is treated with *Dpn*I to eliminate the parental methylated DNA strands. Transformation of the variant DNA strands into *Escherichia coli* can directly repair the nicks and express the variant proteins. This strategy can generate two simultaneous mutations at a distant location.[36,37] The modified method, OmniChange, can generate up to five distant mutation sites simultaneously. However, the primers used in OmniChange need to be specifically synthesized to have 12 bp overlapping tails with phosphorothioate bonds, and additional iodine treatment is required to create the overhangs for annealing purposes.[38] The key drawback of the whole plasmid amplification method is that the undigested parent plasmids always result in high wild-type protein background. PCR-based mutagenesis together with either restriction enzyme based cloning or other elegant DNA assembly protocols is probably the most popular method for preparing single or multisite saturation mutagenesis libraries.[39−41] The key advantage of PCR-based mutagenesis is that multiple primers with one or more degenerate codons can be used to

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

generate multiple distant mutations simultaneously. For example, overlap extension PCR and its modified methods can reliably generate up to six simultaneous mutations.[42,43] The primer extension method that uses phosphorylated primers and T4 ligase together with PCR can generate more than 10 simultaneous mutations;[44] and the "Incorporating Synthetic Nucleotides via Gene reassembly" method can create a library of variants with varying degrees of mutations at different sites.[45]

The key considerations in an SSM experiment are codon degeneracy and SSM strategy. Theoretically speaking, the more degenerate codons used, the more diverse library will be created with increasing probability of finding beneficial mutations. However, using a completely randomized codon (NNN, N = A, C, G or T) results in an extremely large library when mutagenizing multiple sites simultaneously, which makes the screening/selection step very difficult. Thanks to codon redundancy, NNK (K = T or G) reduces the library size by half with 32-fold degeneracy but still encodes the usual 20 amino acids. Further reducing library size can be achieved by using NDT (D = A, G, or T) or DBK (B = C, G, or T) that encode 12 structure-balanced amino acids that result in a library with higher quality in some studies.[46] With more knowledge of protein structure−function relationships, the degeneracy can be further reduced to more specific amino acids that are likely beneficial.[46−48] Step-wise saturation mutagenesis, later called iterative saturation mutagenesis (ISM), is probably the most widely used strategy to create a focused mutagenesis library.[49,50] ISM is based on the Cartesian view of protein structure, performing iterative cycles of saturation mutagenesis at rationally chosen sites that may be crucial for enhancing protein property of interest. Unlike multiple rounds of epPCR, ISM uses the variant gene of a hit from a given library as the starting point for the next round of SSM at the other sites, in which the likelihood of obtaining additive and cooperative effects of newly introduced mutations in the protein space will be maximized. ISM and its derivatives have been widely used to enhance the catalytic properties of enzymes, especially enantioselectivity,[51] and thermostability.[52]

2.1.1.3. Recombination. Homologous recombination plays a key role in natural evolution to recombine beneficial mutations while removing deleterious mutations. Computer simulations demonstrated the importance of iterative homologous recombination in directed evolution because it can explore the sequence space that cannot be accessed by noncombinatorial methods such as epPCR and focused mutagenesis.[53,54] Recombination can occur among single genes with random point mutations[55−58] or naturally occurring homologous genes, which provides "functional diversity".[59,60] This strategy has been proved to be useful in improving enzyme activity,[55,59,61,62] stability,[63−65] and folding,[57,66] and it even can be used to engineer multiple properties of a single protein at the same time, which is hardly achievable by random mutagenesis.[67] DNA shuffling and its derivative methods are popular recombination methods to evolve target genes *in vitro* (Figure 4). DNA shuffling, first reported by Stemmer and co-workers, fragments double-stranded DNAs with DNase I.[68,69] The fragments are randomly reassembled into full-length genes in a primer-free PCR step, resulting in template switching and recombination. The libraries of chimeric genes are amplified by standard PCR and cloned into a vector for further analysis. The main limitations of DNA shuffling are the strong tendency for reconstitution of parental genes and low diversity of the chimeric library since the major crossovers occur at the regions of high sequence identity.[70] To improve the combinatorial efficiency of DNA shuffling, different techniques were developed. For example, "random chimeragenesis on transient templates" (RACHITT) developed by Monticello and co-workers uses single-stranded oligonucleotide (ssDNA) to prepare library fragments and uracil-containing transit ssDNA templates to assist assembly.[71] Instead of using PCR, RACHITT assembles the fragments by hybridizing them with transit templates, trimming the mismatches, filling and ligating the gaps. The heteroduplexed template is then destroyed by uracil-DNA-glycosylase treatment and replaced by a homoduplex chimeric strand during PCR. This method reduces the contamination of parent strands to zero and introduces regions with a sequence identity of 10 or fewer bases at 29% crossover frequency, thus largely improving the crossover resolution. Additionally, combinatorial diversity can also be largely improved by assembling the chimeric genes in yeast.[72]

A few DNA shuffling methods purely based on PCR techniques were developed to simplify the DNA shuffling procedures and improve the crossover frequency (Figure 4). The staggered extension process (StEP), developed by Arnold and co-workers, consists of special PCR cycles with elongation being disrupted earlier due to denaturation.[73] The growing fragment can anneal to the new template in the following cycles and extend further to create chimeric cassettes. The crossover frequency can be maximized by using synthetic shuffling which became popular due to the decreased price of single oligonucleotide synthesis.[67] In this strategy, the diversity from the parent genes is encoded in dozens of degenerate oligonucleotides with homologous overhangs, and the chimeric library is created by primer extension (Figure 4). This method synthetically allows each codon from the templates to combine independently and greatly increases the diversity of chimeric sequences that cannot be achieved by fragment-based shuffling methods. However, it might be challenging for parent genes with low sequence similarity.

The above-mentioned homology-dependent shuffling methods require relatively high levels of DNA homology to recombine genes *in vitro*. However, many proteins sharing similar three-dimensional structures show low or even no perceptible sequence similarities.[74,75] DNA shuffling does not always result in functionally improved protein either. To further expand the application of recombination in protein engineering, a set of homology-independent recombination methods have been developed to explore beneficial combinations of mutations among parental genes with low sequence similarity. Incremental truncation for the creation of hybrid enzymes (ITCHY) creates a fusion library between two genes sharing low sequence similarity by incremental truncation and ligation, which allows a more diverse set of functional fusions, including chimeras with fusion points at nonhomologous sites.[76] ITCHY can also be used together with DNA shuffling to further recombine the library created by ITCHY to access a more diverse sequence space.[77] However, ITCHY generates a large portion of truncated protein fragments because random-sized fragments from two genes are fused to each other. A modified method, sequence homology-independent protein recombination (SHIPREC), fuses two genes head-to-tail before random truncation, and only the fragments with a single gene length are chosen for further analysis.[78] This

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

method largely decreases the number of truncated genes in the final library and improves the efficiency of screening.

Both ITCHY and SHIPREC can generate chimeras with only two crossovers. Nonhomologous random recombination (NRR) developed by Liu and co-workers increases the number of crossovers to as many as 11 per gene.[79] Like DNA shuffling, NRR uses DNase I to prepare a library of random fragments. Differently, the fragments are blunt ended with T4 DNA polymerase and reassembled by ligation. Two hairpin sequences are added in a defined stoichiometry in the ligation reaction to control the final product size that can be further controlled by adding an electrophoresis step. Nonhomologous multicrossover can also be achieved by PCR-amplifying the templates with primers containing Type IIS restriction enzymes. The fragments are digested and assembled by Golden Gate assembly.[80,81] The library obtained from nonhomologous shuffling methods generally contains functional proteins at an extremely low percentage due to mutations, insertions, and deletions. A computational algorithm SCHEMA can be used to predict polypeptide elements that can be exchanged among different proteins with minimum function disruption based on structural information, thus largely decreasing the percentage of nonfunctional proteins.[80,82]

2.1.2. *In Vivo* Gene Diversification. *In vitro* mutagenesis is powerful but includes iterative cycles of gene diversification, transfection, screening, and isolation, which are labor and time intensive. Thus, *in vitro* mutagenesis is generally used to explore the partial sequence space of a single gene. Mutagenesis in intact living cells can avoid repetitive cloning and transformation/transfection steps and can mutate multiple target residues simultaneously. If the desired phenotype can be linked with the cell growth or screened in a high-throughput manner, *in vivo* mutagenesis theoretically can restore and expedite natural evolution in the laboratory. Early applications of an *E. coli* mutator strain (e.g., XL1-red[83−86]) and mutator plasmid (e.g., plasmid encoding *mutD5*[87]) resulted in 4,000- to 5,000-fold greater frequency of single point mutations than the wild-type strain. However, both mutator strain and plasmid have limited applications since they introduce deleterious mutations in the host genome and result in genetic instability. Genome shuffling of microorganisms through protoplast fusion and mating from a diverse population is another popular strategy to engineer Gram-positive bacteria,[88,89] plants,[90] and fungi.[91] It can accelerate directed evolution through genomic recombination without knowledge of the genomic details. However, it is hard to track the genomic changes in the improved variants. In this section, better controlled and tractable *in vivo* mutagenesis strategies will be highlighted.

2.1.2.1. Orthogonal Mutator Plasmids and Mutator Strains. An error-prone orthogonal DNA replication system has been increasingly used in single or multiple gene mutagenesis. In this system, the DNA-dependent RNA polymerase of bacteriophage T7 (T7 RNAP) can specifically recognize the T7 phage promoter. Fusing bacteriophage T7 RNA polymerase with cytidine deaminase can randomly introduce CG → TA mutations on a target gene with a mutation rate of around 10^{-6} in both *E. coli* and human cells.[92−94] The system is very easy to set up and introduces random mutations over a relatively longer sequence (up to 8 kbps) without reducing host viability or inducing "parasite" off-target mutations that lead the cell to circumvent the selection pressure. As an alternative strategy, OrthoRep, a highly error-prone orthogonal DNA polymerase (TP-DNAP1)-DNA plasmid pair, was developed in *Saccharomyces cerevisiae*. It consists of *Kluyveromyces lactis* linear plasmids and their corresponding DNA polymerase.[95] The special replication mechanism of these cytoplasmic plasmids enables the engineered error-prone TP-DNAP1 variants to mutate user-defined genes at a rate of 10^{-5} s.p.b. without increasing the genomic mutation rate ($\sim 10^{-10}$ s.p.b.). However, its complicated experimental procedure limits the application.

B lymphocyte is another "mutator strain" used to engineer antibodies and other proteins. Although genetic information is maintained with high fidelity in most human cell lines, the activated B lymphocytes in the immune system can specifically mutate immunoglobulins (IGs) through somatic hypermutation (SHM). SHM uses an inducible cytidine deaminase and an error-prone DNA repair process to introduce random mutations on IGs at a rate of 10^{-3} m.p.b. while theoretically exerting no mutation load on other genes.[96,97] A human Burkitt lymphoma cell line and a hypermutating chicken B-lymphoma cell line that can mutate IGs and display the encoded immunoglobulin M on cell surfaces can be used to effectively generate antigen-specific monoclonal antibodies.[98] Although SHM has been shown to work on non-IG genes, its application on heterologous proteins is still limited.[99−101]

2.1.2.2. Recombination-Based Methods. Multiplex automated genome engineering (MAGE) is one of the most popular recombination-based *in vivo* mutagenesis strategies in *E. coli*. It was developed based on ssDNA mediated recombineering (Figure 5). Briefly, exogenously introduced

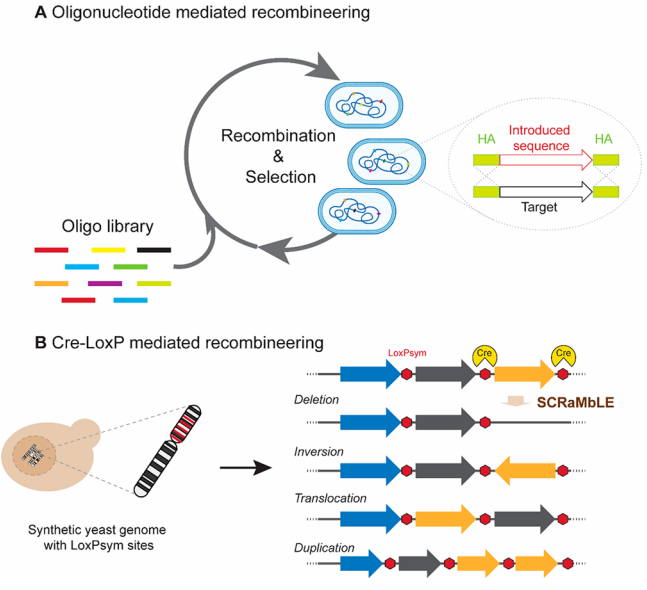


Figure 5. Representative *in vivo* gene recombination strategies in bacteria and yeast. (A) Oligonucleotide mediated recombineering (e.g., MAGE[104]). (B) Cre-LoxP mediated recombineering mutagenesis (e.g., SCRaMbLE[116,117]).

ssDNA can be annealed complementary to the lagging strand at the DNA replication fork by the assistance of the β protein in the bacteriophage λ Red system.[102] The recombination efficiency can be further improved by disrupting the methyl-directed mismatch repair (MMR) system that has been found to correct the base incorporation errors occurring in the progression of the replication fork.[103] MAGE can simultaneously target multiple locations on the chromosome of *E. coli*

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

for combinatorial genomic engineering. Iteratively introducing synthetic DNA allows generation of a huge combinatorial library within a few days for directed evolution.[104]

Due to the inactivation of the MMR system, the accumulation of off-target mutations is one of the main issues in the MAGE system.[105] Instead of turning off the MMR system, overexpression of the *E. coli* Dam methylase can mask the incorporated mismatches from being detected by the MMR system, thus achieving a comparable allelic replacement efficiency but a 33-fold lower off-target mutation rate.[106] The off-target rate can be reduced to the wild-type level by using a precise and portable MAGE system, pORTMAGE, where a dominant negative variant of the *E. coli* MMR protein MutL and λ Red recombinase are expressed under the control of a temperature-sensitive repressor.[107] The MMR system is temporarily turned down to induce the recombination of foreign ssDNAs, while otherwise keeping proper function.[108] Importantly, since the MutL mutator allele cannot be complemented by the native MutL protein,[109] pORTMAGE can be used in different *E. coli* species without tedious strain modification. Low recombination efficiency is another drawback of the early MAGE system (<2% for ssDNA of larger than 20 bps). A coselection strategy (CoS-MAGE) was developed to assist the isolation of highly modified cells by spiking the targeted ssDNA pools with a trace amount of oligos with a selection marker.[110,111] Cos-MAGE increases the recombination frequency from 3% to 15.6% and the frequency of obtaining highly modified clones with multiple mutations more than 200 times.[110] To enable large fragment (several hundred bps) insertion, a synthetic double stranded DNA cassette mediated homologous recombination strategy was developed in the recombination-proficient *E. coli*. The resulting method named trackable multiplex recombineering (TRMR) was used to rapidly regulate more than 95% of the genes in the *E. coli* up or down by integrating functional cassettes in the front of the protein-coding genes.[112]

MAGE has also been adapted to eukaryotes such as yeast and the resulting method is named yeast oligo-mediated genome engineering (YOGE).[114] Exogenous ssDNA-mediated genome editing in yeast is challenging due to the low homology recombination efficiency of ssDNA ($\sim 10^{-4}$).[113] Overexpression of homologous recombination factors Rad51 and Rad54 in an MMR-deficient strain resulted in an enhanced allelic replacement efficiency of 0.1−2%.[114] Interestingly, in another eukaryotic MAGE (eMAGE) system, allelic replacement efficiency was considerably increased (up to 40%) by overexpressing phage λ Red Beta ssDNA annealing protein Rad59 in the Δ*rad51* background strains.[115] This eMAGE system can introduce precise modifications at single base-pair resolution without unintended mutagenic changes at the targeted genetic loci and can simultaneously introduce dozens of mutations (up to 12 ssDNAs with 60 targeted mutations) in one transformation. However, the recombination efficiency dropped tremendously when working with ssDNA with up to 30 bp mismatches or 12 bp insertions.

Another useful yeast combinatorial mutagenesis method is synthetic chromosome rearrangement and modification by LoxP-mediated evolution (SCRaMbLE) (Figure 5), which was established for the synthetic yeast genome project, Sc 2.0. Sc 2.0 aims at synthesizing the entire yeast genome with a built-in diversity generator, such as LoxPsym sequences, that will result in gene deletion, inversion, translocation, and duplication in the presence of Cre recombinase.[116,117] The modifications are first made at the chromosome arm level and then combined. Once complete, more than 5,000 LoxPsym sites will be scattered over the whole synthetic genome. Although Sc2.0 has not been finished yet, SCRaMbLE mediated single synthetic chromosome rearrangement has generated diverse haploids, homozygous, and heterozygous diploids with improved phenotypes, which reveals the great potential of its application in yeast evolution.[118−120] Precise control of SCRaMbLE is critical for creating genomic diversity since continuous LoxP recombination decreases the stability of the synthetic chromosome by deleting some essential genes. SCRaMbLE can be turned on by an β-estradiol inducible system where Cre is fused to the estrogen-binding domain (EBD) of the murine estrogen receptor.[121] Theoretically, Cre is translocated to the nucleus from the cytosol in the presence of estrogen, but its leaky expression results in subtle fitness defects in the synthetic yeast.[119,122] Tighter control can be achieved by using an "AND" gate where expression of Cre-EBD was controlled by the galactose inducible promoter.[119] Cre was only fully expressed when simultaneously adding galactose and estradiol. The activated Cre system can be more easily turned off by using a light controlled split Cre system where each domain of Cre was fused to two heterophytochrome B and its interacting factor PIF3 from the plant. This system also showed lower background recombination.[122]

2.1.2.3. CRISPR-Based Methods. Clustered regularly interspaced short palindromic repeats (CRISPR) are the hallmark of a bacterial defense system where RNA-guided endonucleases (e.g., Cas9, Cpf1) are used to bind and cleave foreign nucleic acids. Various CRISPR-Cas systems have been programmed to use simple 20 bp guide RNAs (gRNAs) to search and target user-defined sequences of different species and edit DNA at precise locations by creating double- or single-strand DNA breaks that are repaired by either nonhomologous end joining or homologous recombination.[123,124] Probably because nonrepaired breaks lead to cell death and cutting continuously occurs until the mismatches disrupt gRNA recognition, CRISPR-Cas improved homology recombination efficiency drastically in both prokaryotes and eukaryotes (from 10^{-5} to ~80%).[125,126] CRISPR-Cas enables large fragment deletion, insertion, and replacement with multiple mismatches, and multiplex editing at high efficiency that cannot be achieved by any previous methods (Figure 6A). For instance, integration of the CRISPR-Cas system into MAGE increased the recombineering efficiency from about 3% to 98% for gene recoding and from about 5% to 70% for small insertions and substitutions.[127] The recombination efficiency remained almost unchanged in *E. coli* with an intact MMR system, which largely reduced the off-target rate associated with the defect mismatch repair system in MAGE.[128]

In yeast, the efficiency of editing three genes simultaneously by CRISPR-Cas system could be as high as 90% by using only 50 bp homology arms (HA).[126] Coupling CRISPR-Cas9 mediated nucleotide replacement with epPCR (CasPER) is a useful tool to introduce random mutations over any 600 bp genomic region with even mutation frequencies.[129] CHAnGE, a CRISPR mediated homology-directed-repair (HDR)-assisted genome-scale engineering method, was developed to precisely knock out each gene in *S. cerevisiae* in a controlled manner.[130] Large-scale engineering is achieved by synthesizing CRISPR guide sequences and the HAs in individual oligonucleotides on microarray chips. Cloning and delivering a pooled CHAnGE plasmid library in *S. cerevisiae* followed by gene editing results

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

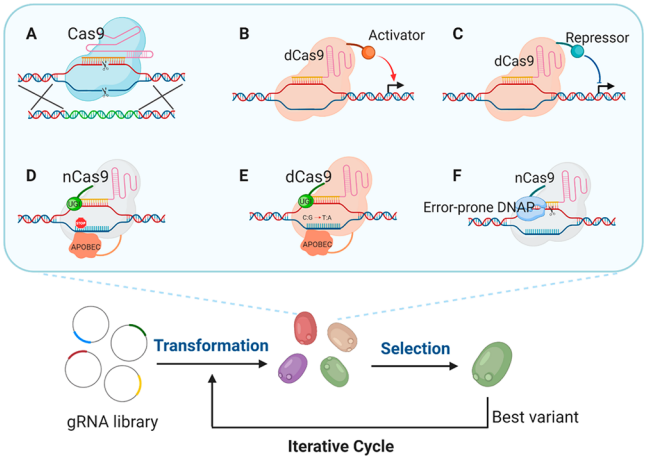


Figure 6. CRISPR-assisted *in vivo* mutagenesis. (A) CRISPR-Cas9-HDR;[125,126] (B) dCas9-mediated gene activation;[134] (C) dCas9-mediated gene repression;[134] (D) genome-wide single gene knockout by base-editing-induced nonsense mutations;[141] (E) random mutagenesis induced by base-editing;[137,138] (F) random mutagenesis induced by nCas9-*E. coli* DNA PolI (error-prone) hybrid protein.[142]

in a yeast variant library for subsequent phenotypic screening. The unique sequence of the gRNA serves as the barcode for variant tracking by next-generating sequencing. This method efficiently edits more than 98% genes with an average frequency of 82% in one transformation and can also be used to perform SSM of the target gene *in vivo*. A few similar CRISPR-Cas9-mediated HDR systems were developed around the same time for both yeast[131,132] and *E. coli*[133] with minor modifications (Table 2).

In addition to controlled genome-wide gene deletion, the multifunctional genome-wide CRISPR (MAGIC) system can precisely overexpress and down-regulate each gene in *S. cerevisiae* at the same time (Figure 6BC).[668] MAGIC was developed based on an orthogonal trifunctional CRISPR system that consists of transcriptional activation, transcriptional interference, and gene deletion in *S. cerevisiae*.[134] It enables the creation of the most comprehensive and diversified genomic libraries by using three independent guide-sequence MAGIC libraries for activation, interference, and deletion, respectively. And it covered more than 99% of all open reading frames plus RNA genes. Additionally, combinatorial gRNA libraries consisting of any two MAGIC libraries enable the discovery of synergistic effects of two genes.

CRISPR-Cas9-mediated HDR can also be utilized to perform directed evolution on proteins in mammalian cells. Using single-stranded DNA oligonucleotides with phosphorothioate bonds at both 3′ and 5′ as homology template can improve the homology repair efficiency up to 35%. Up to nine consecutive degenerate codons flanked by 49 bp single-stranded oligodeoxynucleotide (ssODN) can be used to introduce SSM of 14 a.a. in the target genes. Coupling mammalian cell display and next-generation sequencing, homology directed mutagenesis can be used in a high throughput manner in antibody engineering.[135] Recently, a new HDR strategy was developed to engineer heterologous proteins in the cytosol or subcellular compartments of mammalian cells in a single-variant-per-cell manner. The target protein with up to 59 bp target site deletion was linked with a reporter cassette mTagBFP2. The repairing of the frame-shifted target protein with ssODN containing degenerate codons through CRISPR-Cas-mediated HDR can restore the out-of-frame reporter and edited variants can be enriched efficiently by screening. Additionally, tagging the target gene with lysosome-associated membrane glycoprotein 3 can localize the protein of interest inside the lysosomal lumen for optimizing the protein's function, which is critically dependent on the subcellular compartment.[136]

Modified CRISPR-Cas systems can also introduce random point mutations at target genes without creating double-strand breaks. Using the deactivated Cas9 protein conjugated with an activation-induced cytidine deaminase (AID) variant with partial depletion of its C terminus can introduce random point mutations within a 4 bp window at target loci in human cells with a substitution frequency of around 3% when single gRNA (sgRNA) is used (Figure 6E).[137,138] Although overexpression of AID mainly introduces C to T or G to A mutations in cell lines, dCas9-AIDx can mutate C or G to all the other three bases with an even probability. A random mutagenesis library can be created over a short-range of genomic DNA (150 bps) when a pooled of sgRNA is used. Coexpression of dCas9-AIDx with an uracil glycolysis inhibitor (UGI) increases the mutation frequency up to 20% but restricts the mutations to C → T or G → A substitutions. The mutation window can be expanded to about 20 bps upstream of the PAM sequences when using a sgRNA containing MS2 hairpins in its stem loop which recruits the hyperactive AID variant fused to MS2 binding proteins.[139] However, this strategy results in lower average mutation frequencies, 0.25% to 2.5%, depending on the targeting sites. A similar technique was also developed in plants, where five saturated targeted endogenous mutagenesis editors (STEMEs) were used to mutate endogenous plant genes.[140] Two versions of STEMEs were constructed: STEME-1 was applied to edit cytosine and adenine at the same target site with C > T efficiency up to 61.61% and simultaneous C → T and A → G efficiency up to 15.1%; STEME-NG was used with 20 individual sgRNAs to produce near-saturated mutagenesis at 1 kb gene target[140] Single-nucleotide base substitutions, in-frame indels, and site-specific mutagenesis can be achieved by STEMEs. However, the narrow editing window is still a big issue. In addition to random mutations, base-editors can be used to perform gene knockout by mutating Arg, Gln, or Trp to stop codons (CRISPR-STOP) (Figure 6D).[141] CRISPR-STOP is efficient and less deleterious compared with CRISPR-Cas9-mediated HDR and has great potential for whole-genome knockout experiments in mammalian cells. Instead of dCas9, Cas9 nickase (nCas9) and its variants with broader PAM sequence specificity could create single-strand breaks at the target loci while conjugated cytosine base editors and adenine base editors simultaneously induced C to T and A to G mutations with an efficiency up to 15%.[141] Since the mutation window is also small for each sgRNA, about 12 bps for C → T and 4 bps for A → G, hundreds of sgRNAs are needed to cover a single gene of 1,200 bps.

Coupling nCas9 with error-prone *E. coli* DNA polymerase PolI can also lead to random mutagenesis at target loci in *E. coli* (EvolvR) (Figure 6F).[142] The mutation rate can be tuned by using different PolI variants from about 10^{-7} to 10^{-3}. A higher background mutation rate was also observed for PolI with lower fidelity. Although it was claimed that the editing windows could be up to 350 bps, the majority of the mutations still occurred within 50 bps of the nicking site. Increasing the editing window by using sgRNAs might be critical to improve

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 2. Summary of Gene Diversification Methods

category	methods	principles	advantages	limitations	ref
Mutator strain and plasmid	Mutator *E. coli*	*E. coli* variants with deactivated DNA repair pathways	5,000-fold increased mutation rates	Genetic instability due to uncontrolled, genome wide mutagenesis	83−86
	Mutator plasmid	Curable mutator plasmid carry mutD5 or error-prone Pol I	Lower background mutation; high mutation rate at 10^{-3} s.p.b. (error-prone PoI I)	Mutation occurs near the replication of origin (error-prone PoI I); background mutations	87
	B lymphocytes	Somatic hypermutation in activated B lymphocytes	High mutation rate (10^{-3} s.p.b.) on Igs; minimum background mutations	Mainly for antibody engineering	98−101
	OrthoRep	Error-prone orthogonal linear DNA replication platform in *S. cerevisiae*	High mutation rate (10^{-5} s.p.b.) on target gene; no background mutations	Laborious	95
	Orthogonal T7 system	Polymerase and cytidine deaminase hybrid proteins introducing random mutagenesis during gene transcription	Orthogonal; easy to use; work in both *E. coli* and human cells	Mutation bais: C:G →T:A	92−94
Recombination-based mutagenesis	Genome shuffling	Protoplast fusion for gram positive bacteria; conjugation for *E. coli*; mating for yeast	Convenient and efficient whole genome recombination strategy	Hard to track the relationship between the genotype and the phenotype	88−91
	MAGE	ssDNA-mediated combinatorial optimization of bacterial phenotypes; similar techniques CoS-MAGE, pORTMAGE, YOGE, eMAGE, etc.	Multiplex targeting; work in both *E. coli* and yeast; easy to generate huge combinatorial library	Accumulation of off-target mutation and low recombination efficiency in some cases	104
	TRMR	Double strand DNA cassette-mediated recombination strategies	Up or down regulate multiplex genes in *E. coli*; tractable	N.A.	112
	SCRaMbLE	Cre-LoxP-based recombination on synthetic yeast genome	Can generate diverse haploids, monozygous and heterozygous diploids	Limited to yeast with synthetic genome; lead to unstable genome	116, 117
CRISPR-based mutagenesis	CRISPR-Cas-mediated HDR	CasPER: perform random mutagensis over 600 bp genomic region of yeast	Evenly distributed mutation	Single gene and narrow region	129
		CHAnGE, MAGESTIC, CREATE, etc.: genome wide editing in yeast and *E. coli*	Genome wide gene knockout or SSM on single gene at high editing efficiency	No more than 3-genes could be engineered simultaneously	130−133
		HDR: perform directed evolution on proteins in mammalian cells	SSM on a single gene at high editing efficiency	Single gene and narrow region	135, 136
		MAGIC: genome-wide editing in yeast using orthogonal CRISPR system	Perform gene knockout, overexpression, and knockdown on >99% of all open reading frames and RNA genes	Require high-throughput screening/selection	134, 668
	dCas9-deaminase	CRISRP-AID, CRISPR-STOP, STEME: introduce point mutation on the target site by using dCas9-deaminase conjugation protein	Introduce single bp editing on the target gene in *E. coli*, yeast, human cell and plants	Narrow editing window	137−141
	EvolvR	introduce random mutagenesis on the target site by nCas9-PolI (error-prone) conjugated protein	Tunable mutation rate (10^{-7}−10^{-3} s.p.b.)	Elevated background mutation rate; narrow editing window	142
Retroelement-based mutagenesis	Retrons	Introduce maximum 6 random mutation on the target genomic DNA	Multiple mutation within narrow window	low overall mutation frequency; narrow editing window	146
	ICE	Retroelement-based random mutagenesis of target gene in yeast	high mutation rate (10^{-5} s.p.b.) on the genes without obvious transition/transversion bias over kilobase genes	Challenging to evolve pathways more than 3.5 kbps	147

the efficiency of both dCas9 and nCas9 associated random mutagenesis.

2.1.2.4. Retroelement-Based Methods. Bacterial retroelement "retrons" produce hybrid RNA-ssDNA molecules called msDNA that edit homologous sequences. Retrons contain a guide (msr) and a target cassette (msd) that fold into a secondary structure after transcription.[143] Reverse transcriptase recognizes this structure and reverse-transcribes the msd sequence to form msDNA that can edit homologous genomic regions.[144] Retrons can be programed to express synthetic ssDNAs of interest in *E. coli*, thus acting as a specific genome editing tool.[145] Retrons may edit genes in a fashion similar to ssDNA. Overexpression of λ phage β protein and knocking-out exonuclease genes improved editing efficiency up to 6% and can introduce a maximum of 6 mutations within a 30 bp segment. Transcribing the retron cassette with an error-prone RNA polymerase can introduce random mutations on the target genomic DNA. However, application of this approach is limited due to its narrow editing window and low overall mutation frequency ($\sim 10^{-7}$).[146]

Retroelement-based *in vivo* mutagenesis *in vivo* continuous evolution (ICE), was developed in yeast.[147] The replication cycle of the yeast native retrotransposon Ty1 proceeds through an RNA intermediate that is converted to complementary DNA by reverse transcriptase and integrated into the genomic DNA.[148−150] The error-prone nature of Ty1 replication cycles enables random mutagenesis of heterologous genes inserted between long terminal repeats (LTR) of Ty1 with a mutation rate of up to 10^{-5}.[147] The overall transposition frequency was improved to 10^{-2} by tuning the expression of key regulators of Ty1 transposition, which ensures the practical library size to cover all variants. Impressively, this system can introduce random mutations without obvious transition/transversion bias over kilobase genes and thus can be used to mutate a metabolic pathway.[147]

2.1.3. Computational Tools for Gene Diversifications. Directed evolution experiments are often labor intensive due to

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 3. Representative Examples of Computational Tools for Gene Diversification

category	software	principles	features	ref
Sequence-based	FOLDEF	Analyze importance of the interactions contributing to stability	The FOLDEF is available via a web-interface. It can be used as a fast method for *ab initio* structure prediction.	648
	SIFT	Analyze amino acid substitution significance	SIFT can be used to eliminate functionally neutral substitutions and prioritize substitutions contributing to phenotypic differences.	151
	PANTHER	Analyze amino acid substitution significance	The hidden markov model scores, derived from observing amino acid substitutions, can be used to predict whether a given alleles functionally deleterious.	152
	WebLogo	Generate sequence logo graph	WebLogo is useful to generate graphical representations of sequence patterns for visualization.	649
	PoPMuSiC 2.1	Predict protein stability upon site mutations	PoPMuSiC is available via a web server. It can be used to estimate the optimality of each amino acid in the sequence with respect to structural stability.	650
	PROVEAN	Predict functional effect of amino acid substitutions and indels	PROVEAN is available via web server. It can generate predictions for all types of mutations. It can also support high-throughput analysis.	153
	CAVER 3.0	Analyze transport pathways in dynamic protein structures	CAVER offers a web server for geometry-based analysis of pathways in both static and dynamic protein structures.	156
Structure-based	LPC and CSU software	Analyze interatomic contacts and interface complementarity	The software can be used to calculate the solvent-accessible surface; determine the contacting residue; and indicate all hydrogen bonds.	651
	I-Mutant 2.0	Predict stability change of mutated protein	I Mutant can be used to automatically predict protein stability upon single point mutation with either structure input or structure input.	188
	Q-SiteFinder	Predict ligand binding sites	Q-SiteFinder can be used to identify ligand binding sites for virtual screening and protein residues within a suitable range of the probe clusters.	652
	CASTp	Analyze surface features, functional regions	CASTp offers a web server and can visualize the annotated functional residues, with emphasis on mapping to surface pockets and interior voids.	155
	CUPSAT	Analyze protein stability changes upon point mutations	CUPSAT is available via web server. It can predict variant stability based on existing protein structures or based on uploaded custom structures.	653
	SITEHOUND-web	Predict ligand binding site	SITEHOUND offers a web server that can identify regions of the protein characterized by favorable interactions with a probe molecule, given a protein structure in PDB format.	654
	ConSurf 2010	Analyze evolutionary conservation in sequence and structure	ConSurf is available as a web server. It can calculate the evolutionary conservation of amino acid positions by using either sequence or structure input.	655
	LigPlot+	Analyze ligand-protein interactions	LigPlot+ is a graphical system that can be used to visualize hydrogen bond interaction patterns and hydrophobic contacts.	656

the requirement for iterative rounds of diversification and selection of a large library. Moreover, the efficiency of directed evolution can be greatly influenced by the quality of the library. To address these limitations, a "semi-rational" prescreening can be used to reduce the size of the library and simplify experimental screening. To design a high-quality library which is more likely to yield positive variants, various computational tools have been developed to identify promising target residues. In this section, the computational tools are classified based on the information utilized by the models into sequence-based computational tools, structure-based computational tools, molecular dynamics simulations, and data-driven machine learning tools.

2.1.3.1. Sequence-Based Methods. Sequence-based computational tools typically predict the effects of mutations by performing multiple sequence alignment of homologues of the query protein sequence. The advantage of sequence-based methods is that amino acid sequence is all the input required by the model. They are also relatively computational inexpensive. A few popular methods are introduced below, while additional selected methods are listed in Table 3.

The most widely requested task is prediction of the effect of an amino acid substitution on protein function. Sorting intolerant from tolerant (SIFT), a tool developed by Henikoff and co-workers, can accomplish such a task and requires only a protein sequence as an input. Through aligning the sequences from the protein family, SIFT assesses whether the substitution of an amino acid residue is deleterious or not by evaluating if the position is conserved evolutionarily. SIFT can predict if a substitution has the potential to alter the phenotype based on the sequence similarity between the query sequence and its homologues in the database.[151] Aiming to develop a comprehensive and curated database of protein families and functions, Thomas and co-workers developed protein analysis through evolutionary relationships (PANTHER), which is capable of relating protein sequence to function. PANTHER consists of two parts: library and index. The PANTHER library represents protein families and subfamilies as a multiple sequence alignment, a hidden Markov model, and a family tree. The PANTHER index is an abbreviated ontology. PANTHER was used to predict protein function on a database scale.[152] Another widely used method named PROVEAN (Protein Variation Effect Analyzer) can predict functional change not only from single amino acid substitutions but also other sequence variations including multiple substitutions and in-frame insertions and deletions.[153]

Notably, the major advantages of sequence-based tools are the easy-to-access input and low-cost computational power. Most of the tools have web-based servers with user-friendly interfaces. These methods are customizable by users and can generate results relatively fast.[154]

2.1.3.2. Structure-Based Methods. Structure-based computational tools are especially useful when identifying hot spots for protein engineering. These methods require the 3D structure of the query protein as input. Many of the structure-based tools make predictions by analyzing interatomic contacts, surface features, substrate interactions, and many other physical properties.

Liang and co-workers developed a computed atlas of surface topography of proteins (CASTp), an online software, that can locate and measure pockets and voids in 3D protein structures.[155] CASTp uses sequence alignment methods to obtain sequence mapping between annotated residues in Swiss-Prot, OMIM and the structures in PDB. The web server can be used to study surface features, functional regions, and the function of key residues. CastP provides a detailed analysis of

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 4. Representative Examples of Machine Learning-Assisted Protein Engineering Studies

application	data	embedding method	ML model	ref
Thermostability	2,087 single point mutations	pH, temperature, residue, and spatial environment information encoded	SVM	186
	3,366 single and multisite mutations	Structural feature vector embedding	SVM	657
	2,156 single point, 180 double, and 141 multiple point mutation	Structural feature vector embedding	Random Forest	183
	ProTherm data set	Structural feature vector embedding	Multiple	184
	1,925 data set, SKEMPI data set and ProNIT data set	Graph-based atom distance patterns	Gaussian Processes	192
	242 P450 mutation data set	Structural residue−residue contact map	Gaussian Processes	169
	ProTherm data set	Conformation of residue, spatial distance between two amino acids, and solvent accessibility	Neural Network	199
	ProTherm data set	Residue interaction networks	Neural Network	658
Solubility	58,689 soluble and 70,954 insoluble protein sequences	Feature representations learned by a deep learning model	Convolutional Neural Network	200
Enantioselectivity	145 variants	Residue−residue interactions	SVM	659
Membrane localization	218 variatns	One-hot and structural feature vectors	Gaussian process	173
Binding affinity	PDBBind data set	Distance matrix and neighbor list	Neural Network	660
	IEDB data set	Word embedding	Recurrent Neural Network	661
Substrate for enzymatic reactions	BRENDA database	K-mer vector representations	Gaussian process	193
Secondary structure	6,128 proteins	Orthogonal encoding and sequence profiles	Recurrent Neural Network	662
	6,128 proteins	One-hot	Generative stochastic network	663

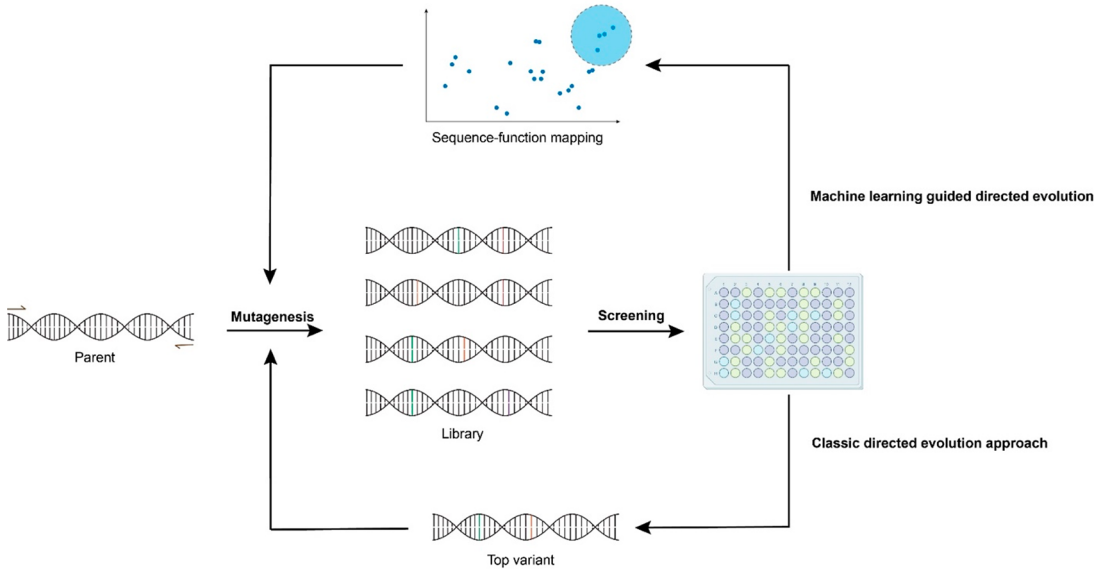


Figure 7. Workflow comparison between classic directed evolution approach and machine learning guided directed evolution approach.

the pockets of the protein and is thus most useful for studies involving binding pockets. Strand and co-workers developed a software named CAVER 3.0 that can identify and characterize the transport tunnels and channels in macromolecular structures. Taking protein structural information as an input, CAVER 3.0 analyzes the tunnels and channels in large ensembles of protein conformations and can facilitate the study of molecular transport, recognition, and enzyme catalysis.[156] Similar to sequence-based tools, some structure-based tools incorporate additional structure information to analyze the mutational effects of amino acid sequences. HotSpot Wizard, initialized by Damborsky and co-workers, is a web server for identifying hot spots for protein engineering. The server integrates multiple computational tools and provides users with the predicted hot spots as well as functionally important residues.[157]

2.1.3.3. Molecular Dynamics. The key idea of a molecular dynamics (MD) simulation is simple: the simulation calculates the interactions between individual atoms within a biomolecular system over time. Thus, the simulation can be used to predict the position of each atom at any given time by iteratively updating the future position and velocity based on current information. MD simulation is a powerful technique because the simulation environment is highly controlled and the simulation can capture the position and velocity information for each atom in the system which cannot be experimentally obtained.[158] The libraries generated by "semi-rational" approaches can produce higher hit rates compared with full coverage libraries.[159] Such approaches are beneficial

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

for identifying hot spots as well as reducing library sizes. MD simulation-assisted directed evolution methods have been used to engineer proteins with improved catalytic activity,[160−162] enantioselectivity,[163−165] stability,[166] specificity,[167] and thermostability.[168]

2.1.3.4. Machine Learning. The screening of a large variant library is often a key bottleneck in directed evolution. Even after multiple iterations of a high-throughput selection or screening method, only a minimal subset of the full sequence landscape can be explored.[169−171] With the rapid development of data science, machine learning has been increasingly applied to protein engineering. Its main aim is to build an accurate sequence-to-function landscape from measured data and guide experiments to explore the landscape more efficiently. Some selected examples are shown in Table 4. Different from an experimental approach, a machine learning-assisted directed evolution workflow aims to build a sequence-to-function mapping based on quantified data from screening the variant pool. An optimization algorithm is followed by landscape construction and locates the top performing variants in the landscape, as shown in Figure 7. The top performing variants will then be used for further iterations of learning and screening. To establish this workflow, three key steps are needed, including (1) representing protein sequences as vectors, (2) building a sequence-function model, and (3) exploring the model and locating optimum variants.

Before building a sequence-to-function mapping landscape, protein sequences need to be first encoded into vectors. Different data representations have varying levels of integration of the information hidden in the protein sequence and lead to varied learning performance.[172] One of the most intuitive ways to encode a protein is one-hot encoding, wherein a protein of length L is encoded into a matrix with shape L by the number of amino acids. The element *i,j* in the matrix is one if the corresponding *i*th position of the sequence is the *j*th amino acid, and zero otherwise. The density of information stored in one-hot encoding is relatively low. However, one-hot encoding can often be used as a baseline method.[173] With the rapid advances in natural language processing (NLP) in the computer science field, highly efficient NLP methods can be employed for the protein embedding task. For example, language models are trained based on valid protein sequences and predict the most probable amino acids given the previous amino acid sequence.[174,175] Because unlabeled protein sequence data is readily available, unsupervised models are often used to train a representation model. Such methods vectorize a string of sequence into a fixed length continuous vector. Doc2Vec, for example, converts various lengths of sequences into fixed length vectors so that the similarity of the sequences is directly reflected by the distance between the vectors.[176] Long short-term memory neural network (LSTM) is a robust NLP method and has accomplished numerous tasks.[177] A protein-embedding method named UniRep (unified representation) takes advantage of LSTM and extracts the hidden states as the unified representation of a protein. Such embedding methods are able to distill a wide variety of information such as secondary structure and stability from sequence information alone.[177] Recently, the most successful and widely used model for NLP is the transformer model.[178] For protein embedding, the transformer model was proven to outperform many previous state-of-the-art models for certain tasks, such as stability prediction.[174] Employing an unsupervised learning scheme, more sophisticated models can find patterns from a large amount of sequence data and distill the information in the form of a fixed length vector. Such methods usually have rather high informational density but require tuning according to homology information to further increase the accuracy of the target protein.

The second major step is to choose a suitable machine learning model to capture the sequence-to-function landscape. Different machine learning algorithms can be implemented depending on various factors such as number of data points available, computational resources, and desired accuracy.[171] One of the simplest machine learning models is a linear model, which applies a linear transformation to the input data set.[179,180] Linear models have a small number of parameters and can be used when the number of available data sets is limited, in which case, complex machine learning models with more parameters can lead to overfitting. Combined with a good protein embedding method, linear models can make accurate predictions about function.[181] A slightly more complicated model is a decision tree model, which uses a tree structure to map an input to a classification by going through multiple queries which are represented by the branches. Random forests is a common algorithm which is an ensemble of weak prediction models that, together, form an accurate prediction model.[182] Decision tree models, such as random forests, have been used to accurately predict enzyme stability.[183,184] The Kernel method is a pattern analysis algorithm that avoids explicitly mapping input to typically a higher dimensional feature space but takes advantage of kernel, which captures the similarity between two input data points and maps the raw input to the feature space implicitly.[185] Support vector machine (SVM) is one of the best-known kernel methods and has been applied to predict protein stability and enantioselectivity.[186−189] Gaussian process is a probability-based predictor, which uses a kernel method to predict an unobserved data point from training data.[190] Unlike the other algorithms mentioned above, Gaussian process does not predict the value of the unobserved data but captures uncertainty by Gaussian distribution. Gaussian process has been used to predict protein stability[169,191,192] and activity.[193,194] On the complicated side, neural networks map an input layer to an output layer by training the parameters in the hidden layers embedded between the input and output layers. Neural networks can be accurate because of their high complexity but can potentially cause overfitting when there is not enough training data. Neural networks are used in numerous protein engineering tasks including prediction of property,[195−198] stability,[199,200] and structure.[201] Generative model is a newly evolved method for biosystem design applications. Under the scheme of unsupervised learning, generative models learn the distribution of the training data set and generate data points from the distribution with some variation.[202] EVmutation, for example, captures the covariation between all residue pairs and predicts new variation by quantifying the change of one or more residues.[203] DeepSequence maps protein sequence to a latent space which captures higher order dependencies between residues and then recovers new variants from the latent space.[204] The training of generative models often requires a large amount of homologous sequences to achieve a high accuracy but little to no mutagenesis data due to its unsupervised nature.

The last major step of machine learning guided directed evolution is to locate the optimum based on a well-trained machine learning landscape which reflects the sequence-to-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

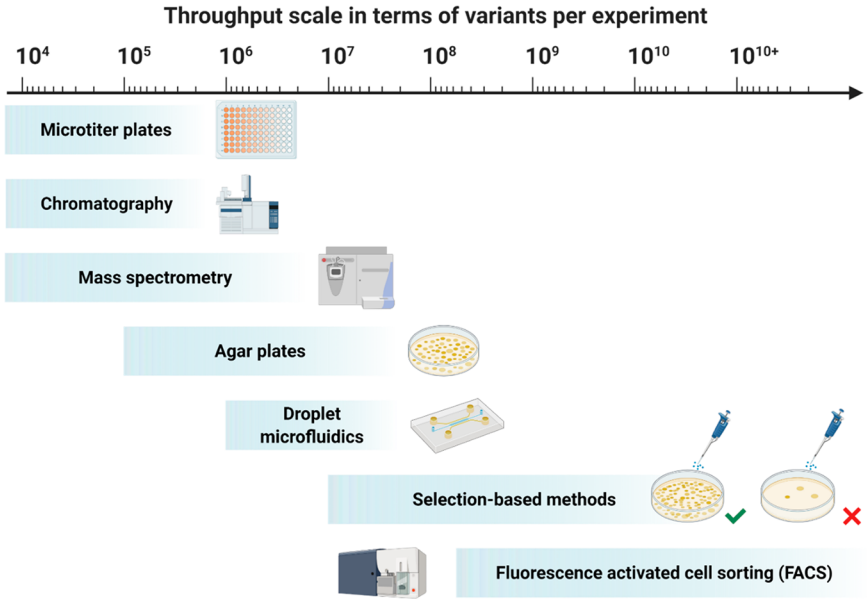


Figure 8. Throughput-scale spectrum of the potential size of variants per experiment for different analysis methods.

function mapping. One way to explore the landscape is using Bayesian optimization, which guides the choice of unobserved data point exploration to maximize the exploration efficiency.[205] Such methods have been used to navigate protein fitness landscapes for stability engineering.[169] Markov Chain Monte Carlo simulations have also been used to simulate directed evolution *in silico* to fast forward the discovery of optima as demonstrated by the engineering of TEM-1 β lactamase and green florescent protein.[181] Recently, a predictor algorithm was designed to robustly locate top performing positive variants from an enormous variant landscape using a generative model.[206]

In summary, a wide variety of machine learning algorithms have been used for directed evolution purposes. However, there is no single algorithm that suits all protein engineering applications across the board.[207] One of the major factors in choosing an algorithm is the number of available training data points. If plenty of data is available, a complex model such as a neural network may yield the highest accuracy. If the data are limited and susceptible to overfitting, a simple model such as the linear or kernel method may give the best result. The desired accuracy is also an important factor, in such case, using a simple model as a baseline and evaluating its accuracy before moving into more complicated models is worthy.

2.2. Selection and Screening

Rapid advances in library creation technologies such as mutagenesis, recombination, and advanced computational tools have made it possible to generate large gene libraries processing more than a billion variants in a short period of time. As mentioned before, one key challenge in directed evolution is to build a general platform for identifying desired phenotypes efficiently and analyzing target products in a high-throughput manner. Currently, library analysis can be categorized into either selection-based or screening-based methods. For selection, the protein function needs to be linked to the growth or survival of the host organism. The nonfunctional variants will be automatically eliminated during the selection process. This method can evaluate a large library containing more than 10^9 variants and only proteins with the desired properties will be distinguished (Figure 8). For screening, each library variant is assayed individually by using biochemical or biophysical analytical methods to evaluate the desired property. This method is versatile and flexible because the experimental conditions can be easily tailored to meet a specific industrial setting, such as a non-natural environment or substrates, and the analytical tools are generally applicable for a range of enzymatic transformations. However, the throughput of the most screening-based techniques is in the range of 10^4 to 10^6 variants per experiment, which is low and still a major limitation (Figure 8). In sections 2.2.1 and 2.2.2, we provide a detailed overview of selection and high-throughput screening methods for diverse applications, including enzyme/protein engineering, pathway engineering, and genome engineering. General principles, significance, novelty, and specific examples are summarized for each method.

2.2.1. Selection-Based Methods. By imposing certain selective pressures on the protein libraries, selection-based methods can directly eliminate undesired variants and subsequently allow positive candidates to be subjected to subsequent rounds of directed evolution. This feature enables selection-based methods to assess a larger library than screening-based methods. Currently, a variety of high-throughput selection methods have been developed for selecting proteins with different binding affinities and catalytic activities, peptides that can bind target molecules and regulate their functions, and RNA molecules that bind specific ligands. In general, selection methods can be divided into four major groups: display-based selection, compartmentalization, biopanning with phage-displayed peptides, and systematic evolution of ligands by exponential enrichment (SELEX) (Table 5).

2.2.1.1. Display-Based Selection. Display methods connect a protein to the DNA encoding it. Such powerful tools have found widespread applications in selecting binding complexes with higher affinities against various target molecules in the directed evolution of protein-based binders.[208] The way to display proteins on the surface of an organism depends on the physical linkage between the DNA molecule and the protein it encodes. If there is a direct linkage, it is plasmid display, ribosome display, mRNA display, or SNAP display (Figure

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 5. Summary of Selection/Screening-Based Methods for Library Analysis, including Their Advantages, Disadvantages, And Applications/Targets

category	methods for library analysis		advantages	limitations	applications/targets	ref
Selection-based methods	Display-based selection	Phage display	Display multiple copies of proteins; library diversity	No generic phage available; no PTM-based mechanisms; fusion proteins	Substrate specificity, binding affinity, product stability, activity	208−213
		Retrovirus display	Allow PTMs; smooth folding of eukaryotic proteins	Fusion proteins expression	Infection of mammalian cells for protein expression	214, 215
		Plasmid display	High stability; no issues related to protein secretion or *in vitro* translation	Fusion proteins expression	Protein binders	224
		Ribosome display mRNA display	High efficiency; large library; no compatibility issues between *in vitro* transcription and translation	No PTM-based mechanisms; fusion protein expression; mRNA is not stable	Protein binders; catalytic enzymes; high affinity antibodies	225−228
		SNAP and SNAP dendrimer display	Display multiple copies of proteins; selections under harsh conditions	No PTM-based mechanisms; weak binders are not stable during display	Protein binders	229, 230
	In vivo compartmentalization	Growth complementation	Select enzymes with diverse properties; high sensitivity and efficiency	Growth conditions vary among different microbes; complex genetic regulation networks; high error rates	Enzyme properties relate to the fitness of the host microbe	231, 232, 664
		Chemical complementation			Modifications of substrates link to reporters	233−235
		Genetic complementation			Two-hybrid systems	241−247
	Biopanning		High-throughput selection of peptides interactions with other substances	Achieve strong affinity only after multiple iterative rounds of biopanning	Affinity-based selection for peptides binding to individual targets	252−270
	SELEX		Large library; diverse targets of interests; high affinities and specificities; adapt various conditions during the selection	Without specific modifications for different targets, a standardized protocol for aptamer development is not available	Nucleic acid ligands/aptamers	271−274
Screening-based methods	FACS-based screening	FACS for cell surface display	Display and express enzymes directly on the outer surface of cell; avoid cell lysis	Fusion proteins expression	Sorting of bacteria, yeast, and mammalian cells; binding assays to engineer receptors	217−222
		FACS for *in vitro* compartmentalization	High sensitivity and efficiency; avoid cloning steps; large library	No PTM-based mechanisms; cannot screen enzymes with compatibility issues between *in vitro* transcription and translation	Enzymatic activities link to FACS or fluorescence	248−251
		FACS for reporter	Select enzymes with diverse properties	Each reporter is independent and specific; hard to be applied industrially	Enzymes with transcription regulation molecules	281−283
	Microtiter plates		Applicable to many assays	Time-consuming and labor intensive	Enzyme reactions with colorimetric or fluorometric assays; cell culturing in flasks or bioreactors	284−287
	MS-based screening	Electron impact ion sources on GC; ESI sources with LC, CE, direct infusion; desorption/ionization with MALDI, SIMS	Permit label-free analysis of enzyme reactions and protein−ligand interactions with high specificity, sensitivity, and efficiency	Require special instrumentations, highly trained personnel, and bioinformatic pipelines; complicated data sets and time-consuming sample preparation steps	Protease; kinase; PTM enzymes; carbohydrate-active enzymes; acetylcholinesterase; ligands and inhibitors for biomedical, therapeutic applications	288−297
	Droplet microfluidics	Mass spectrometry	Label-free, high sensitivity, enable detailed analysis of the chemical structures	Require pretreatment; matrix effects; sometimes require ionization	Chemical synthesis; molecular biology, biomedicine; food science; environmental science	299−302,308
		Optical detection	High sensitivity; easy to read	Mostly need fluorescence labeling; cannot be applied to improve stereoselectivity; limited to substrates with fluorophores or chromophores	Enzymatic activities link to FACS or fluorescence	303, 305
		Electrical detection	High sensitivity, rapid response, low cost	Need electric-active molecule	Electrochemistry, impedance, or capacitance	306
		NMR spectroscopy	Chemical structure analysis; label-free	Low sensitivity and throughput	Chemical synthesis and metabolic compounds	307
	Digital imaging		High sensitivity and high-quality data	Limited applications	Colorimetric activity assays	310−312

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

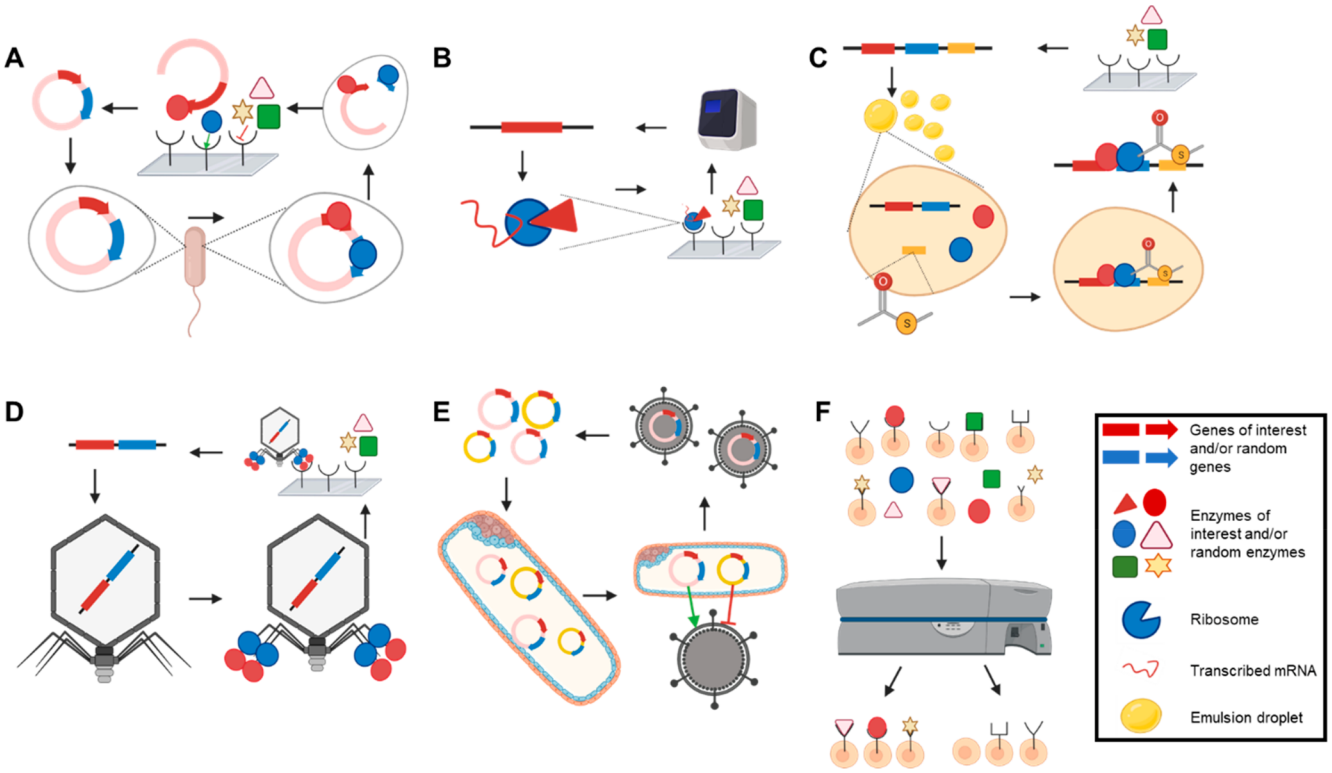


Figure 9. Schematic overview of six display-based selection methods. (A) Plasmid display. The gene encoding the enzyme of interest is transformed. After binding, the encoded enzyme is expressed. The remaining genes can go through another round of selection.[224] (B) Ribosome/mRNA display. The gene is transcribed and translated *in vitro*. The transcribed mRNA and expressed enzyme are associated with the ribosome to form a complex. The selected gene is recovered by real-time PCR.[225−228] (C) SNAP display. The gene is expressed in an emulsion droplet. The encoded enzymes are covalently bound to DNA through a thioester bond and form a complex for extraction and selection.[229] (D) Phage display. The gene is transformed into a phage particle. After expression, the encoded enzyme is displayed on the phage surface followed by selection.[4] (E) Retrovirus display. The plasmid libraries are integrated into mammalian cells with only selected plasmids transfected into virus.[214] (F) Cell surface display. Enzymes are displayed on the cell surface and bind to fluorescent molecules. After cells pass through a sorting machine, candidates are separated into various populations.[217]

9A−C). Otherwise, it is phage display, retrovirus display, or cell surface display (Figure 9D−F). For direct display methods, the upper limit of throughput mainly depends on the amount of DNA produced in each experiment; for indirect display methods, the throughput depends on the transformation efficiency of the host cells.

As the first display method, phage display has been well studied over the years. To enable the enrichment of desired variants, multiple rounds of iterative selection to infect bacteria by phage is necessary. Phage display has been successfully utilized for altering substrate specificity, increasing the binding affinity of a protein, and engineering enzymes with improved stability and activity.[208−213] However, because each phage display system requires specific knowledge of a particular protein of interest, there is no generic solution available for improving multiple properties of different targets simultaneously. In addition, expression of some enzymes in eukaryotes often requires proper folding and post-translational modifications (PTMs), but phages lack such mechanisms. To circumvent these limitations, retrovirus display was developed and has been applied to infect mammalian cells for protein expression.[214] For instance, Merten and co-workers showed a more than 1,300-fold enrichment of active wild-type tissue plasminogen activator during a single selection cycle by using retrovirus display.[215] In addition, displaying proteins on the surface of cells has been widely developed for binding assays to engineer receptors such as antibodies and T-cells with high affinity.[208] This method can be applied to enzymes with PTMs as well. Many different scaffold proteins have been used in the cell surface display for enzyme engineering. Since enzymes with encoded DNA can be expressed and displayed directly on the outer surface of the cell, they can be distinguished and separated easily via a cell sorting/counting-based instrument.[216] Therefore, display technologies are often coupled with fluorescence-activated cell sorting (FACS) for high-throughput screening.[217−222] For example, by fusing the target to the C-terminus of the Aga2 cell surface agglutinin protein, yeast display allows exposure to fluorescently labeled substrates in the media, which enables this display to be coupled with FACS for high-throughput screening.[219,220,223]

For plasmid display, a DNA-binding protein is noncovalently fused to the encoded protein. The fusion protein is then expressed in the cell and binds to its encoding plasmid. After the cells are lysed, the protein−plasmid complex can be selected according to different recognition DNA sequences. Yoo and co-workers demonstrated the ability to discover functional proteins from large libraries via plasmid display. The GAL4 DNA binding domain was constructed to enrich the molecular diversity for *in vitro* selection of target proteins from a protein mixture.[224] Ribosome display links the protein of interest to the corresponding genetic codes by creating a complex of mRNA, ribosome, and translated protein. Multiple pools of sequences enriched for the desired binding affinity can be obtained after exposing the complex to the protein's binding

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

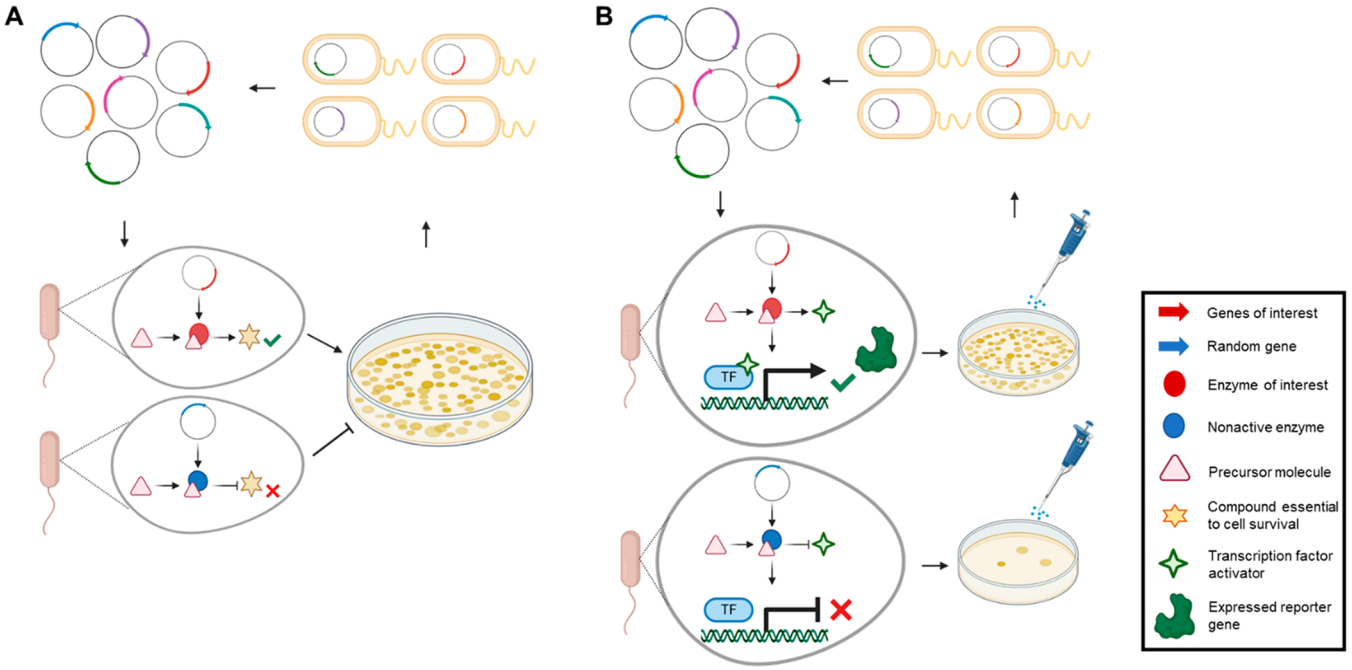


Figure 10. Schematic overview of *in vivo* compartmentalization. (A) Genetic/growth complementation. A library of variant plasmids is transformed into host cells. The gene encoding the enzyme of interest binds a precursor molecule and converts it into a compound that is essential to cell survival, while the other nonactive genes are not able to generate this essential compound. After plating the cells onto selection plates that lack the compound essential to cell survival, only cells with active enzymes survive and are selected for the next round of mutagenesis.[231,232] (B) Chemical/reporter complementation. The gene encoding the enzyme of interest binds a precursor molecule and converts it into a transcription factor (TF) activator, which binds to the TF of the corresponding gene. The TF dissociates, and the reporter gene is expressed, while the other nonactive genes and enzymes are not able to generate this transcription activator. After plating the cells onto plates with specific antibiotics, only cells with active enzymes survive and form colonies. The active genes are recovered, amplified, and subjected to the next round of mutagenesis.[233−235]

partner and washing away nonbinding complexes.[225] Similarly, mRNA display produces proteins that are covalently bonded to their encoding mRNA by *in vitro* transcription and then linking with the peptidyl acceptor antibiotic puromycin.[226−228] Since there are no transformation steps involved during library creation, the speed of selection for mRNA and ribosome display can be quick. Besides, the size of the library can reach up to ~10^{13} variants per experiment for these *in vitro* selection methods. Another display technology called SNAP display allows each phenotype to be covalently bonded with its encoding genotype through the SNAP-tag. The complex is then encapsulated in water-in-oil emulsion droplets where *in vitro* DNA transcription and translation take place.[229] For SNAP display, the physical linkage of genotype and phenotype is a covalent bond, which allows selection to occur under harsh conditions such as high pH or temperature. The library size of this type of display is up to 10^{9} variants per 1 mL of emulsion.[229] In one study, an improved SNAP display system was constructed after three rounds of selection to achieve a 10^{7}-fold enrichment of a Her2 binder over nonbinding proteins.[229] Built on SNAP display, SNAP dendrimers display enables multiple copies of a protein of interest to be mounted on a single DNA scaffold encoding these proteins. It takes advantage of avidity effects during affinity panning and is a useful tool for *in vitro* directed evolution of protein binders.[230] Hollfelder and co-workers showed that multivalent SNAP dendrimers can enhance the enrichment of candidate binders by up to 5-fold higher and recovery by up to 25-fold higher when compared to the normal SNAP monomer display method.[230]

2.2.1.2. Compartmentalization. In compartmentalization, the movement of proteins and genes is arbitrary yet restricted in a three-dimensional containment area. Compartmentalization can be divided into *in vivo* compartmentalization and *in vitro* compartmentalization according to the nature of the single compartment used in a specific protein-gene interaction. Phage particles, bacteria, yeasts, and even mammalian cells are categorized as cell-like *in vivo* biological compartments. Water-in-oil emulsion droplets and water-in-oil-in-water double emulsion droplets belong to man-made *in vitro* compartmentalization. They can encapsulate only DNA and the essential transcriptional-translational machinery to encode the protein of interest.

For *in vivo* compartmentalization, various chemical, genetic, and growth complementations have been implemented to the cell-like compartments acting as the specific selection conditions. With DNA libraries introduced as a ligation mixture or plasmids, the cell itself provides a segregated compartment and acts as an expression host. Moreover, thanks to the optimization of transformation techniques for plasmids or ligation mixtures, such as electroporation and chemical transformation, single DNA variants can be introduced into each cell with a maximal efficiency of ~10^{10} clones per μg DNA. Cellular assays are also highly suited for selection when the property of interest is linked to a type of complementation that can show differences/changes among DNA variants phenotypically. For example, yeast auxotrophic strains can be used to separate variants with disrupted essential metabolic functions from normal strains by monitoring cell growth rates. Antibiotics like ampicillin, kanamycin, or penicillin have also been supplemented to bacteria or mammalian cells in rounds

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

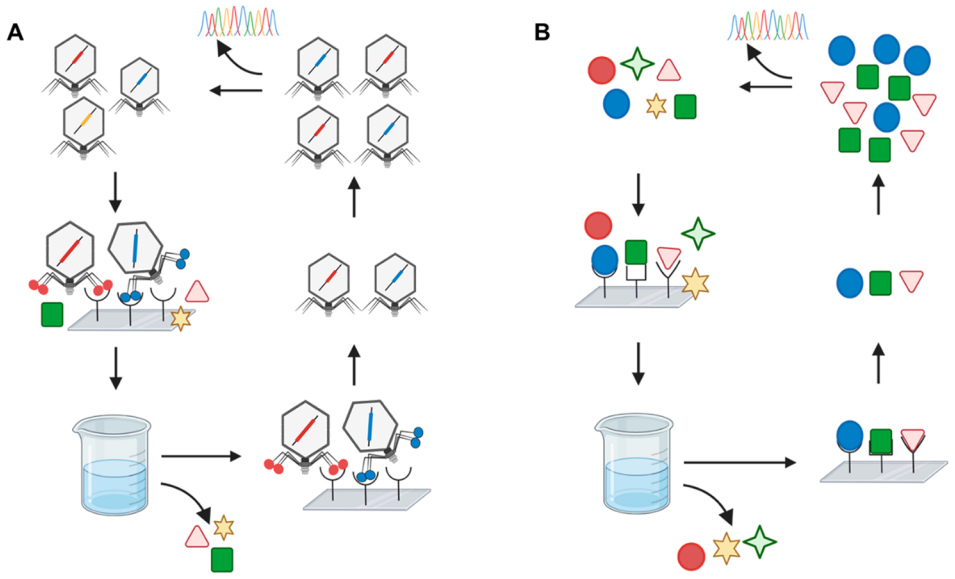


Figure 11. (A) Biopanning workflow. A gene library is inserted into a region of the bacteriophage genome. The peptide product is displayed on the surface of a virion which can bind the desired target to form a complex. Peptide-target binding affinity is utilized as a selection criterion: phages with weak affinity are washed away from the displayed surface and phages with strong affinity are kept. Next, bound phages are eluted and enriched through infection and amplification in bacterial cells. After several rounds of selection panning, the best candidates are collected and sequenced.[252] (B) Systematic evolution of ligands by exponential enrichment (SELEX). SELEX begins with binding between the DNA or RNA library and the target molecules. Next, strongly bound candidates are kept and unbound oligos are removed by extensive washing of the complexes. The target bound oligos are eluted and amplified by PCR for DNA or real-time PCR for RNA. The selected oligos with high binding affinities to the target molecules are sequenced and enriched for the next round of evolution.[271−273]

of directed evolution to improve the resistance and survival rate of certain proteins under harsh environmental conditions.

For *in vivo* genetic and growth complementation, the activity being investigated is intrinsically linked to the fitness of the host microbe because only the cells containing the desired enzyme variants with specific genetic modifications can grow under the selective pressure (Figure 10A). For example, overexpression libraries were constructed inside genetically modified *E. coli* K-12 strains that were unable to utilize glucose as a carbon source. Proteins with latent glucokinase activity were selected after rounds of growth.[231] Other enzyme properties such as enantioselectivity can also be efficiently distinguished via growth complementation. A dual selection system was developed to select for a lipase with desired enantioselectivity in one study.[232] After three rounds of selection with increasing phosphonate ester concentration, a variant with an improved enantioselectivity toward the *S*-enantiomer was selected.

In vivo chemical complementation is a more general approach than *in vivo* genetic and growth complementation since phenotypical changes do not have to happen directly on the specific gene or enzyme being studied. In chemical complementation, one can modify the substrate of the targeted reaction and link the outcome with an unrelated reporter system (Figure 10B). For instance, a variant transcriptional activator was used to bind various benzoic acids to detect the activity of benzaldehyde dehydrogenase.[233] The well-studied reporters, e.g., β-galactosidase or amino acid−based selection markers, have also been used as phenotypic readouts due to color changes or growth rate differences of cells across variants.[234,235] Theoretically, *in vivo* selection is able to assess a large number of variants rapidly because only cells harboring variants with functional improvements can grow into colonies. However, due to the complex genetic regulation networks of microorganisms and the lack of related enzyme information, this approach has met limited success.

Overall, cell-like *in vivo* compartmentalization has been used to increase expression levels of soluble proteins,[236,237] resistance to antibiotics,[238] or the ability to metabolize certain carbon sources.[239,240] With the development of two-hybrid systems in protein engineering, *in vivo* compartmentalization has been applied for the selection of protein−DNA, protein−RNA, and protein−-small molecule interactions with improved catalytic activities.[241−247] However, false-positive rates can be high due to the generation of background mutations in host cells. Furthermore, *in vivo* compartmentalization cannot be used to identify the desired phenotype under extreme temperatures or pH conditions. In such cases, additional selection/screening-based techniques such as visually identifiable fluorescence-based markers or colorimetry may become necessary to implement and support the phenotypical characterization.

On the other hand, man-made *in vitro* compartmentalization links genotype and phenotype by coupling transcription and translation within a water-in-oil or water-in-oil-in-water emulsion system.[248] Genes with associated proteins, substrates, and products are all contained within 2 μm diameter aqueous droplets for the selection of various catalytic properties in parallel.[248] As genes are transcribed and translated *in vitro*, cloning procedures can be avoided to expedite the selection, and the size of library is only limited by the amount of DNA used rather than the transformation efficiency. Moreover, most *in vitro* compartmentalization strategies have been coupled to FACS for analyzing very complicated enzyme properties. For example, the activity of homocysteine thiolactonase or β-galactosidase can be detected by FACS because a fluorescent signal is generated in the substrate or product of interest.[249−251]

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

2.2.1.3. Biopanning with Phage-Displayed Peptides. Phage display is one of the most popular methods for studying protein−protein, protein−peptide, and protein−DNA interactions. Very large libraries of variants on a scale of $\sim 10^{10}$ per experiment can be expressed in *E. coli* and then displayed on active phages by inserting the target protein into one of the coat proteins of filamentous bacteriophage (pVIII for high copy display, pIII for low copy display).[211,213] Random phage-displayed peptide libraries provide a physical linkage between the displayed peptide and the encoded DNA, which enables one to quickly identify peptides that can bind target molecules and regulate their functions after phenotypic selection. Use of affinity selection to identify peptides binding to individual targets is termed biopanning. There are four major steps involved in biopanning for peptide selection (Figure 11A). The first step is to create a phage display library with foreign genes of interest inserted into a region of the bacteriophage genome. The peptide product will then be displayed on the surface of a virion. Second, the constructed phage library is bound to the desired target to form a phage-displayed peptide library complex. The next step is to wash away the unbound phages with weak affinity from the displayed surface. Only bound phages with strong affinity are kept. Lastly, bound phages are eluted using high concentrations of acids and salts or variations of pH. Binders can subsequently be enriched through infection and amplification in bacterial host cells.[252] Multiple rounds of biopanning are recommended so that peptides with strong binding affinity to the target can be obtained.

In summary, peptides that bind to specific target molecules can be readily isolated by implementing iterative rounds of the biopanning with the phage-displayed peptide libraries. In the past two decades, this phage-display technology platform has been used for different applications such as B-cell and T-cell epitope mapping,[253−255] selection of disease-specific antigen mimics,[256,257] selection of organ-specific peptides,[258−262] selection of cell-specific peptides[263−266] and ligands,[267−269] and selection of peptides bound to nonprotein targets.[270]

2.2.1.4. SELEX. SELEX is an *in vitro* selection method that aims to evolve nucleic acid ligands, also known as aptamers, for new functionalities.[271,272] When an oligonucleotide, such as RNA or ssDNA, binds to its target with high selectivity and sensitivity, a specific and complex three-dimensional shaped aptamer is formed. Due to the flexibility of their structures, aptamers can be generated to bind a wide variety of targets ranging from inorganic and small organic molecules, peptides, proteins, to complex mixtures and even whole cells. Aptamers have been widely used for basic pharmaceutical research, drug discovery, and investigation of binding phenomena in proteomics. SELEX has been modified over the years to help with selection of aptamers with high affinities efficiently and quickly. A large library of $\sim 10^{15}$ different oligonucleotides can be selected in parallel through SELEX. Ligands with the highest target binding affinities will be easily amplified and filtered out to facilitate the *in vitro* selection.[273]

The basic steps of *in vitro* selection of target-specific aptamers using SELEX are summarized in Figure 11B. First, a synthetic random ssDNA or RNA oligonucleotide library with a size of 10^{13}−10^{15} different sequence motifs is created. Second, binding between the DNA or RNA library and the target molecules occurs, which can separate strongly bound, weakly bound, and unbound oligonucleotides. Unbound oligonucleotides are removed by extensive washing of the complexes. Third, the target bound oligonucleotides are eluted and amplified by PCR for DNA or reverse transcription (RT)-PCR for RNA. Selected oligos with high binding affinities for the target molecules are enriched in a pool for the next round of selection. Similar to directed evolution, SELEX mimics Darwinian evolution where repetitive cycles of *in vitro* selection and enzymatic amplification can drive the process toward a motif with the highest affinity and specificity for the target.[274] The number of rounds of SELEX depends on multiple parameters, such as target molecule features, selection environment, or partitioning efficiency. In general, six to 20 SELEX rounds are needed to obtain aptamers with high affinity and specificity.[273] Finally, the last SELEX cycle occurs when the desired motifs are achieved. In this last round, the enriched pool of the best aptamers will be cloned after the amplification step. The individual aptamers can be sequenced for further characterization or subjected to different post-SELEX modifications and binding studies.

SELEX is a powerful and easy-to-use tool with many applications. It enables efficient selection of evolved aptamers which bind a variety of target molecules in a specific and high affinity manner. Evolved aptamers have been used in therapeutic applications, especially as therapeutic agents that can compete with antibodies. For example, the first aptamer-based therapeutic agent named pegaptanib (Macugen) was used to treat age-related macular degeneration by targeting vascular endothelial growth factor.[274−277]

2.2.2. Screening-Based Methods. In comparison to selection, screening-based methods evaluate each individual variant. Since the coverage of candidates is high, this strategy greatly reduces the chance of missing a desired variant. However, the scale of analysis and throughput become a trade-off. Typically, variants in a library (with a size of $\sim 10^5$ per experiment) can be screened and analyzed individually with high accuracy. Several of the most frequently used screening methods will be described below (Table 5).

2.2.2.1. FACS-Based Screening. Based upon the specific light scattering and fluorescence characteristics of each sample, FACS can sort a mixture of biological cells individually into two or more containers (Figure 12). It can provide efficient and specific recording of fluorescent signals from each cell as well as separate cells of interest during flow cytometry. Over the years, FACS has become a frequently used tool for sorting individual cells of interest.[278,279] In addition, it has been coupled with many other screening/selection approaches for analyzing different enzymes at a rate of up to 10^5 cells per second in a quantitative manner.[217,279,280]

The FACS process involves several steps: (1) The sample containing a population of cells is injected in the middle of a fluid sheath. This sheath is achieved by having a tube with the inner walls built up of a moving stream of fluid. Once the cells are injected into the middle of this sheath, a stable two-layer flowing fluid will form since the two liquids will not mix due to their difference in viscosity. (2) The cells are then hydrodynamically focused inside the funnel as a straight line. (3) The cells pass through a laser beam that measures the forward scatter channel (FSC) for information on the cell size and side scatter channel (SSC) for information on the fluorescence and type of cell. (4) The cells pass through a nozzle that vibrates and results in cell droplets with positive or negative charge. The positively charged cells are attracted to the negative side of the electromagnet, while the negatively charged cells are attracted to the positive side of the electromagnet. (5) A charged droplet containing the cell of interest is directed by an

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

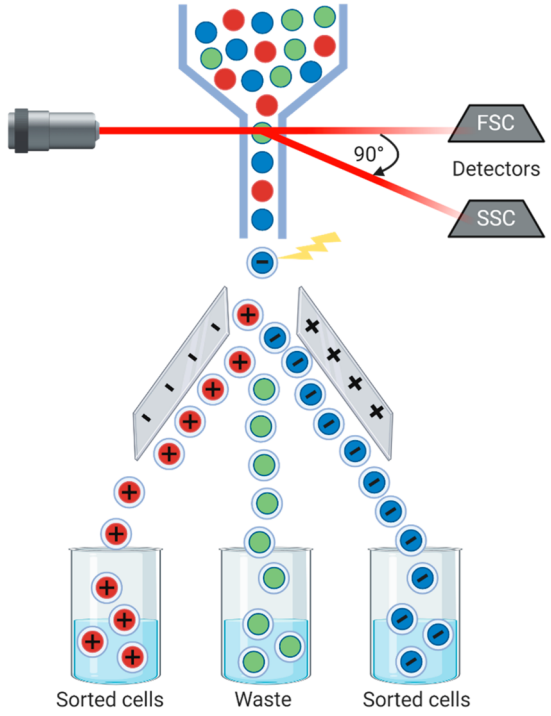


Figure 12. Schematic overview of the fluorescence activated cell sorting (FACS) process. FSC: forward scatter channel. SSC: side scatter channel.

electromagnetic field into the corresponding collection tube (sorted cells or waste). In many studies, green fluorescent protein (GFP) is an ideal reporter for cell fluorescence analysis.[281,282] Riboswitches or ribozymes have also been used, which are turned on upon binding a specific small molecule. Michener and Smolke introduced a synthetic riboswitch coupled with FACS to aid in the directed evolution of a caffeine demethylase.[283]

2.2.2.2. Microtiter Plate-Based Screening. Microtiter plates containing 96 to 9,600 wells have been combined with colorimetric or fluorometric assays to screen for various enzyme properties. In a microtiter plate assay, the protein with its encoding DNA is compartmentalized in a single well; hence, enzymes with different activity profiles are easily distinguished and unable to diffuse or mix with other DNA−protein pairs.[284−286] The substrates and products can be easily identified by macroscopic observation or measuring UV/vis absorbance or fluorescence through a plate reader (Figure 13).[286] Recently, a microbioreactor system named Biolector has been used to detect protease and cellulase activities.[284,285] Biolector enables online monitoring of the light scattering intensity and reduced nicotinamide adenine dinucleotide (NADH) fluorescence signals by adding light excitation and emission filters to the microtiter plate on a specially designed continuous shaking device. The scattered light signal difference between hydrolysis of an insoluble protein substrate and NADH-coupled enzyme indicates the protease and cellulase activity levels.[284,285] Microtiter plates have also been used as an alternative to shake flasks for growing variants before screening.[287] However, without the help of an automated colony picker and liquid handler, the throughput of the microtiter plate-based method is relatively low with a screening library size of $\sim 10^4$ per experiment.

2.2.2.3. Mass Spectrometry (MS)-Based Screening. Recent advances in high-throughput MS resulted in highly sensitive and accurate methods for analyzing complex analytes, which have been especially useful for protein engineering applications.[288−292] By utilizing native substrates and ligands, MS is capable of performing label-free high-throughput assays for characterizing complex mixtures of engineered proteins. However, no individual MS technique provides the ability to screen all proteins in a sample with enough information to characterize all possible modifications. Therefore, it is important to choose the appropriate MS-based approaches.

In a typical MS-based screening method, assay creation, sample preparation, MS measurement, and data analysis are necessary steps. Each step can be performed in a variety of formats, which depends on the nature of the target protein, information to be obtained, throughput, and instruments. First, assays can be set up either in solution or on surfaces (Figure 14a−c). For solution-based assays, the previously discussed microtiter plates are commonly used. For surface-based assays, microarrays are widely used, where matrix solutions can be applied to a solid support before screening.[293] Second, sample preparation is critical to limit the number of molecules competing for ionization and thus reduce the background from irrelevant materials. Various strategies can be employed to separate target analytes from complex mixtures. Currently, gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) are the commonly used separation techniques (Figure 14d−g). In most cases, sample preparation remains time- and labor-consuming and hence

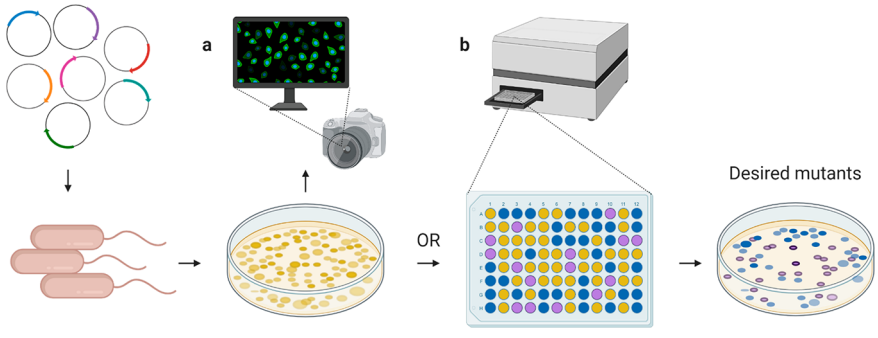


Figure 13. Digital imaging- and microtiter plate-based screening. (a) Digital imaging. A library of variant plasmids is transformed into host cells and plated. Advanced imaging device is applied to individual clones for screening.[310−312] (b) Microtiter plates. A library of variant plasmids is transformed and plated. Enzyme variants are expressed inside the cells. The cells are lysed, and the lysates are transferred into a microtiter plate for enzymatic assay. The enzyme activity can be visualized by macroscopic observation or plate reader.[284−286]

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

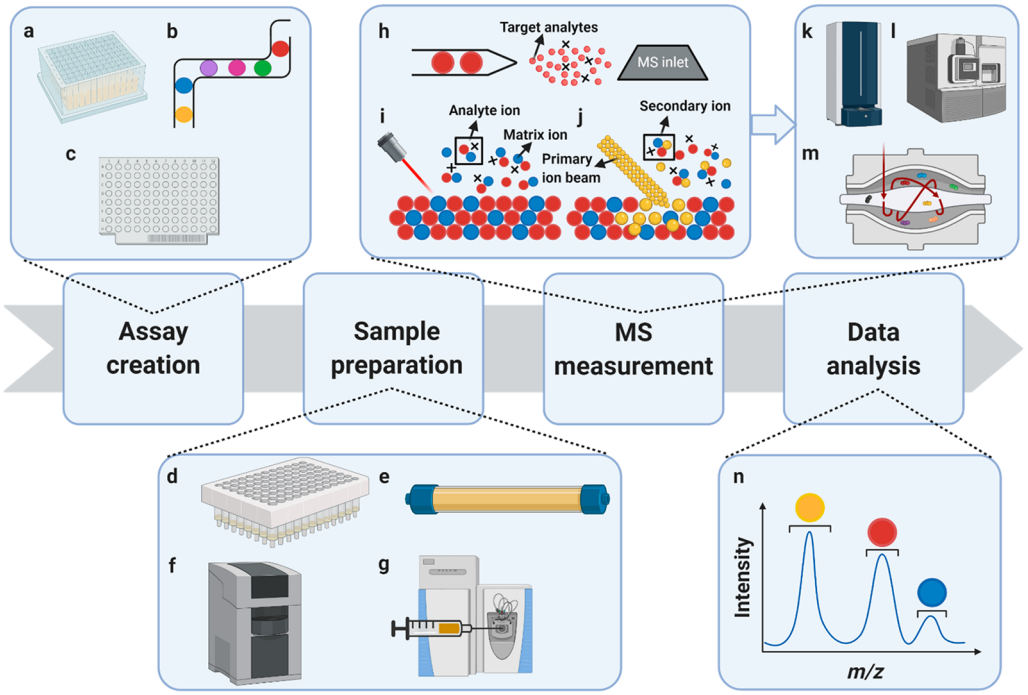


Figure 14. High-throughput mass spectrometry (MS)-based screening.[288−297] (a) Solution assays in microtiter plates. (b) Segmented flow of microfluidic droplets. (c) Assays on microarray surfaces. (d) Solid-phase extraction for separating target analytes from complex mixtures. (e) Chromatographic separation column. (f) Capillary electrophoresis-based sample separation. (g) Direct sample infusion into a mass analyzer. (h−j) Ion sources. (k−m) Ion analyzers. (h) Electrospray-based ionization. Target analytes are transferred inside a spray needle containing microfluidic droplets. At high voltage, these analytes are ionized into single charged droplets and moved into MS inlet for mass analysis. (i) MALDI-based ionization. After the high voltage laser beam shoots onto the surface of sample containing target analytes and matrices, analytes are ionized, and the charged analyte-matrix complex forms. (j) Secondary ion mass spectrometry (SIMS)-based ionization. After the primary ion beam shoots onto the surface of sample containing target analytes, matrices, and primary ions, analytes are ionized, and the charged secondary ions forms. (k) MALDI-ToF mass analyzer. (l) Triple quadrupole-based mass analyzer. (m) Orbitrap mass analyzer. (n) Data analysis. The analyte ions are separated based on mass-to-charge (m/z) ratios in the mass analyzer, and then quantified by the ion detector.

becomes the rate-limiting step in MS-based assays. Third, to initiate an MS measurement, samples need to be introduced into the ion source of a mass spectrometer. Gas-phase ions are then generated from neutral molecules in the ion source, separated based on mass-to-charge (m/z) ratios in the mass analyzer, and quantified by the ion detector (Figure 14h−m). Next, relative ion abundance is plotted versus m/z values to generate a mass spectrum, which contains either qualitative or quantitative information for the measured analytes (Figure 14n). Mass spectrometers are usually categorized based on ion sources and mass analyzers. For example, electron impact ion sources are often used on GC. Desorption/ionization processes are frequently coupled with matrix-assisted laser desorption/ionization (MALDI).[294] The mass analyzer determines the detection limit, mass resolution, and quantitation capability of an MS-based platform. The time-of-flight (ToF) mass analyzer has been widely used due to its relatively low cost, large m/z detection window, label-free feature, and fast scan rates. Detection limits on the zeptomole to attomole range are achieved with ToF MS while maintaining a mass resolution above 20,000 dots per inch.[292,295,296]

As an example of high-throughput MS-based screening for directed evolution, Zhao and co-workers developed an optically guided MALDI-ToF MS to engineer multistep enzymatic reactions through high-throughput, direct screening of microbial colonies.[294,297] This method was used to characterize the substrate tolerance of a five-enzyme pathway synthesizing the antibiotic plantazolicin from a ribosomal precursor peptide and to screen for rhamnolipid congener compositions produced by a two-enzyme pathway after directed evolution. Improved MS acquisition efficiency and information-rich insights were obtained by this technique on large populations of colonies at a rate of 1−2.5 s per colony.[294,297]

Overall, MS-based assays have been mostly used for screening compound libraries against a select protein target, primarily aiming at ligand and inhibitor discovery for biomedical applications. However, since evaluation of individual variants is normally required in MS-based screening for directed evolution, and sample preparation takes time, increasing the throughput of MS-based methods still faces a lot of hurdles.

2.2.2.4. Droplet Microfluidics. Droplet-based microfluidics has emerged as an attractive platform for high-throughput screening that requires very small volumes of analytes. Miniaturized protein assays in femto- to nanoliter-scale plugs or droplets can be generated in capillary or microfluidic channels.[298,299] Consequently, operations such as reagent addition, dilution, splitting, and sorting are performed through microfluidic manipulations. By encapsulating individual variants inside microfluidic droplets to perform enzymatic reactions, the throughput can reach up to 10^8 samples per hour.[15] Next, for rapidly displaying the readouts, these plugs or droplet reactors are coupled to different types of analytical detection techniques including MS platforms such as electro-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

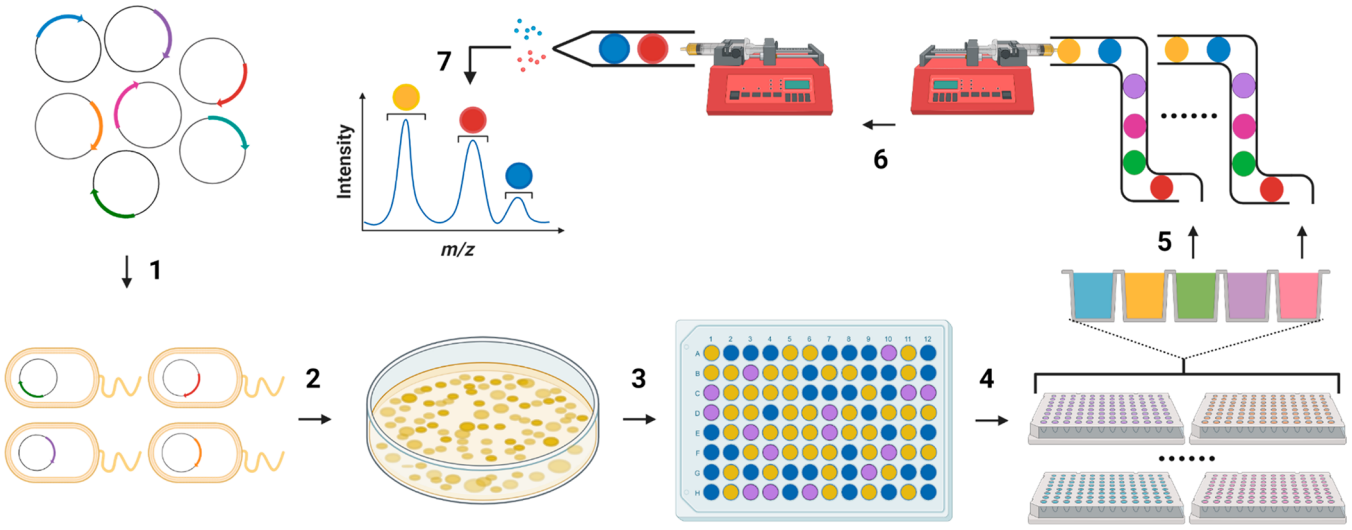


Figure 15. Droplet microfluidics.[299−301] Following the arrows: (1, 2) DNA-carrying plasmids are created and transformed into host cells; (3) the cells are cultured, and the target enzymes are produced after cell lysis; (4) enzymatic assays are dispensed into microtiter plates; (5−7) oil-segmented droplets are generated and directly infused into mass spectrometry source for analysis.

spray ionization (ESI)[300,301] or MALDI;[302] optical detection methods such as FACS,[303] absorbance,[304] or Raman,[305] electrical detection;[306] and nuclear magnetic resonance (NMR) spectroscopy.[307]

The general workflow for droplet microfluidics involves several steps: (1) DNA templates of selected enzymes are created on plasmids; (2) the DNA-carrying plasmids are transformed into host cells such as *E. coli* for amplification; (3) the cells are cultured, and the target enzymes are produced; (4) the cells are lysed to release the enzymes and the enzymatic assays are dispensed into multiple microtiter plates; (5) sample droplets are generated inside Teflon tubes in parallel from the enzymatic assay plates; and (6) these droplets are segmented by perfluorodecalin in a tube and pumped through microfluidic channels. Enzyme and quencher are added sequentially via microfluidic manipulation. Reaction mixtures can be detected via optical techniques in general, but this requires a color or fluorescence change of the reaction being studied. Alternatively, sample mixtures can be infused into a metal-coated, fused silica ESI or CE emitter for MS analysis: (7) best performing candidates from data analysis can then proceed with another round of directed evolution (Figure 15). Such droplet microfluidics settings have been applied in a variety of disciplines ranging from ligand/inhibitor screening to protein engineering.[299,308] In addition, droplet microfluidics also enables the discovery and engineering of rare targets due to the feature of high screening capacity. Hollfelder and co-workers identified new hydrolases with promiscuous activities after screening a million-membered metagenomic library in microfluidic picoliter droplet compartments.[309]

Currently, analytical detection techniques that use fluorescence or absorbance signals as the readout are still the method of choice for droplet-based high-throughput assays.[299] Moreover, the enzyme assays developed for the droplet sorter are limited to substrates with fluorophores or chromophores. Consequently, such droplet technique can hardly be applied industrially when aiming to improve enzyme selectivity. However, the continuous development and evolution of other advanced detection approaches such as MS and NMR platforms will tremendously help the wide adoption of droplet-based high-throughput assays in chemical and biological laboratories.

2.2.2.5. Digital Imaging. Digital imaging combines single-pixel imaging spectroscopy with a solid-phase screening method to screen enzyme variants with high efficiency (Figure 13). Since this technology relies on simple, widely known colorimetric activity assays, it can be applied to a broad range of targets and allows screening of problematic substrates in solid phase or highly viscous solutions that are normally hard to characterize by traditional high-throughput screening methods.[310−312]

2.3. Continuous Evolution

Continuous evolution aims to enable gene diversification, expression, and screening or selection to occur iteratively without human intervention. It can shorten the experimental time and increase the total number of rounds of evolution, thus largely enhancing evolution effectiveness in an evolutionary search.[17,313] Continuous evolution can be performed either in *vitro* or *in vivo*. *In vitro* continuous evolution has been mainly used to evolve RNA ligase ribozymes.[314] It requires the ribozyme to ligate a chimeric DNA-RNA substrate that has the sequence of the promoter element for a DNA-dependent RNA polymerase.[314,315] After reverse transcription, cDNAs derived from an active ribozyme containing a functional promoter element will be generated. Multiple copies of RNA per copy of DNA template can be obtained after repeated cycles with reverse transcriptase and RNA polymerase.[314,315] Population size, sequence diversity, and selection pressure can all be controlled. The rapid pace of continuous *in vitro* evolution makes it powerful and allows one to perform tens to hundreds of generations per day.[314,315] In addition, since the workflow of this method is easy to handle, only minimal human intervention is required. However, it has limitations because several criteria must be met for ribozyme-catalyzed reactions to occur. For example, a particular oligonucleotide must attach to the 5′ end of the ribozyme to enable reverse transcription. Besides, reaction conditions, such as pH and temperature, need to be compatible with the reverse transcriptase and RNA polymerase.[315] The various applications of continuous *in vitro* evolution can be found in section 3.1.1.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Several *in vivo* mutagenesis methods mentioned in section 2.1.2 can be used continuously to evolve a target protein whose product can be associated with the cell growth under certain selection pressure, such as mutator plasmid-based methods,[316] T7 RNAP-based methods,[92,93] EvolvR,[142] retroelement based methods,[145,147] and inducible SCRaMbLE.[119] The primary challenge of this strategy is the co-occurrence of adaptive evolution that can create "cheaters" whose target proteins do not contribute to the improved phenotype. If the protein of interest cannot be correlated to cell growth, *in vivo* mutagenesis methods are generally used to generate a neutral drift library, followed by screening/selection. Except for the above-mentioned *in vivo* mutagenesis methods, several breakthroughs in continuous evolution have been made during the last 10 years.

Phage-assisted continuous evolution (PACE) is one of the most successful continuous *in vivo* evolution methods. It can execute more than 30 rounds of protein evolution over 24 h without human intervention, which can hardly be achieved by other existing directed evolution methods. Instead of linking the activity of target proteins to host cell growth, which can easily result in background mutations from adaptive evolution, PACE links the target protein's activity to the production of infectious progeny phage.[317] To achieve this goal, the essential gene III, encoding protein III for F pilus binding and host cell entry, is deleted from the phage vector and inserted into an accessory plasmid (AP) expressed in the *E. coli* host cells. An arabinose-inducible mutator plasmid (MP) encoding *dnaQ926*, *recA730*, *umuD′*, and *umuC* is installed in *E. coli* to increase the error rate during DNA replication by suppressing the proofreading and enhancing error-prone lesion bypass, respectively.[318,319] The evolutionary procedure is set up in a chemotactic lagoon where the *E. coli* host cells continuously flow through and are infected with selection plasmid (SP) encoding a library of variants (Figure 16). Since the concentration of *E. coli* host cells and progeny are constant, only SP encoding variants that induce a higher level of pIII production will infect more *E. coli* cells and persist in the lagoon, while nonfunctional or low activity SP variants do not. Since the *E. coli* cells only stay in the lagoon for about 0.3 h and are nonproliferating, the neutral drift variants do not accumulate in the *E. coli* cells. To make the PACE method more flexible, a stringency modulation module is applied to adjust the selection pressure using tetracycline, which enables evolving proteins to obtain drastically changed properties.[320,321] In addition, a negative selection module, in which a negative variant of gene III is controlled by a theophylline-activated riboswitch, is available for evolving proteins with highly specific new activity.[320]

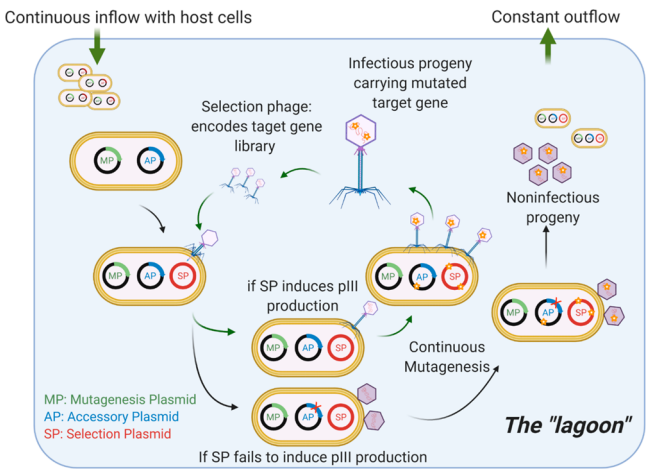


Figure 16. Schematic of phage-assisted continuous evolution (PACE).[317] (Adapted with permission from ref 317. Copyright 2011 Springer Nature. Original Figure 1 was adapted in this figure.)

PACE can theoretically evolve any proteins that can be linked to the production of pIII, e.g., DNA-, RNA-, and peptide-binding proteins, proteases, and aminoacyl-tRNA synthetases. The design for directed evolution is straightforward in PACE. For example, AP with pIII expressed under the specific promoter can be used to improve RNAP activity and specificity.[317,320,322] Other DNA binding enzymes, such as Cas proteins and transcription activator-like effector nucleases (TALENs), can be fused to the ω subunit of bacterial RNA polymerase III.[323] The binding of this fusion protein to user-defined sequences upstream of a minimal *lac* promoter can induce transcription of the downstream pIII, enabling the enzyme to evolve with enhanced specificity.[324,325] Protein–protein interactions can also be evolved by modifying a bacterial two-hybrid system correspondingly. For example, fusing the protein of interest (e.g., *Bacillus thuringienesis* toxins and antibodies) with the ω subunit of RNAP III and its target protein to HA4 monobody that can bind to the SH2 domain of ABL1 kinase at the upstream of pIII gene links the interactions between the protein of interest and the target protein with progeny proliferation rate.[326,327] Protein interaction between smaller units such as biosensors can be evolved using a split T7 RNAP component where two units of a biosensor are conjugated with two split domains of RNAP. The activated sensor will bring two RNAP domains together for pIII transcription.[321] Additionally, PACE can also evolve the reactivity and specificity of enzymes such as proteases and aminoacyl-tRNA synthetases (AARS).[328−330] The protease substrate can work as a cleavable substrate linker connecting T7 RNAP and its inhibitor T7 lysozyme. Gene III can only be expressed when the protease cuts the linker, which restores the active form of T7 RNAP.[328,329] Two strategies can be used to select active AARS out of the variants pool. The amber stop codon can be either introduced in T7 RNAP at permissive sites or in gene III.[330] Only AARS-catalyzed aminoacylation of an amber suppressor tRNA enables the transcription of full-length RNAP or gene III.

A PACE-like method has also been established in human cells.[331] A custom-designed adenovirus variant without essential adenoviral DNA polymerase (AdPol) and protease (AdProt) but containing the gene encoding the protein of interest serves as the "cargo" to accumulate the functional variants. An error-prone AdPol together with AdProt are constitutively expressed in human cells. AdPol enables gene diversification during virus replication; and AdProt is expressed depending on the activity of protein. Only viruses with functional variants can propagate and accumulate in the cell culture during serial dilution. The evolution stringency can also be adjusted by using a small-molecule adenoviral protease inhibitor. Although this method is not wholly continuous since manual culture transfer is required, it is a useful platform to evolve proteins in human cells. It can work in different cell types, mutate large genes (up to 7 kbps), and potentially evolve multiple proteins simultaneously.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Although PACE has been successful in evolving several proteins, specific genetic circuits or elements need to be designed for the different proteins of interest. PACE cannot be used to evolve a protein whose activity cannot be linked to the expression of a reporter protein, and it also cannot be used to evolve multiple enzymes or a biochemical pathway simultaneously. More sophisticated automated cultivation systems need to be designed to continuously evolve cells by using the *in vivo* mutagenesis methods discussed in section 2.1.2. For example, a MAGE automation device has been set up to combinatorially evolve up to 50 different genomes simultaneously.[104] One MAGE cycle consists of seven different experimental steps including growing cells at 30 °C; inducing replacement of the β protein expression oligo at 42 °C; chilling the cells to 4 °C, followed by a cell wash for competent cell preparation; adding and delivering oligos by electroporation; and recovering the cells in media. The completed cycle can be executed by a temperature-controlled growth chamber with real-time cell density monitors; a syringe pump system for buffer exchange, cell transferring and concentration; and real-time generation of competent cells for electroporation with synthetic DNA. This system does not include a selection/screening unit; iterative cycles are run to generate a large library with enough cell diversity for phenotype screening. However, it is possible to integrate specific selection units for cell growth-related phenotypic screening.

Continuous *in vivo* evolution based on cell growth is also challenging since traditional culturing methods either lack precise control or are hard to perform in a high-throughput manner.[332−334] Therefore, precise tuning of selection pressure is essential to guide the populations evolving through the desired paths efficiently.[326,329,335] Ideally, the continuous instant feedback of mutation fitness under the current selection pressure will automatically adjust the selection condition. Without precise control and adjustment, the beneficial mutations are easily "washed out" of the system if the selection stringency is outpaced by the adaption rate. Thus, trial-and-error experiments are needed to identify suitable selection conditions.[95] Additionally, parallel evolution experiments with different selection strategies are also beneficial to navigate the evolutionary landscape fully. Although some automated cell growth systems can control cell growth precisely under adjustable conditions, they are hard to perform parallelly. eVOLVER is a new continuous culturing system that allows users to continuously grow and monitor hundreds of individual cultures in parallel under user-defined growth parameters.[336] This system consists of three main components: (1) a customizable "smart sleeve" housing and interfacing with individual culture vessels allows the cell growing under defined conditions, such as cell density, temperature, and light intensity; (2) a millifluidic module controls the movement of liquids while preventing the formation of biofilms; and (3) a modular hardware infrastructure controls the functionality of the whole system. The real-time, feedback-controlled tuning of selection parameters makes this system a powerful automation platform that can be potentially coupled with *in vivo* mutagenesis methods to evolve proteins of interest efficiently.[337]

3. APPLICATIONS

3.1. Nucleic Acids

In this section, we focus on the applications of directed evolution for engineering functional nucleic acids.[338,339] In particular, RNA is ideal for directed evolution because it combines genotype and phenotype in a single molecule.[339] The discovery of novel catalytic RNA enzymes (ribozymes) inside large populations of random sequences has initiated many potential applications in gene therapy or the detection of small molecules.[315,339−342] On the other hand, a variety of DNA enzymes can catalyze the cleavage of RNA, and hence inactivate target RNAs in cells or whole organisms. Such RNA-cleaving DNA enzymes have many applications in biology and medicine.[315] Below we give an overview of selected examples including RNA ligase ribozymes, RNA-cleaving DNA enzymes, aptamers, and riboswitches.

3.1.1. RNA Ligase Ribozymes. The first ribozyme-based RNA ligase experiment was conducted by Joyce and co-workers to achieve approximately 300 successive rounds of catalysis and selective amplification in 52 h via a simple serial transfer procedure.[314] As a result, substantial improvement of the catalytic efficiency and amplification rate of the RNAs was achieved during the evolution process. After rounds of mutagenesis, ribozymes with a 1.5 min doubling time and high copy number were isolated from the pool.[314] By examining the sequence and kinetic properties of the improved ribozymes, the group further investigated the ability to evolve a ribozyme's biochemical properties in response to the behavior of another macromolecule, such as a DNA enzyme.[343] Consequently, more applications of ribozymes with enhanced robustness have appeared. In another study, RNA ligase ribozymes that could operate under extreme pH conditions were obtained after continuous *in vitro* evolution.[344] Iterative rounds of selective amplification under gradually more acidic or more alkaline conditions were applied to the ribozymes. The two final evolved populations of ribozymes were able to operate at either pH 5.8 or pH 9.8, respectively.[344] At these two extreme conditions, the low-pH ribozyme exhibited a 10-fold increase in catalytic rate compared to the parent ribozyme; the high-pH ribozyme retained its structural integrity and activity, whereas the parent ribozyme was denatured and had no detectable activity.[344] These findings showed the utility of continuous *in vitro* evolution for exploring enzymatic function under stringent environmental conditions.[345] Continuous *in vitro* evolution has also been used to address the classic question of recurrence in evolutionary biology.[346] When a living system repeatedly evolves from a given starting point and converges to the same end point multiple times, it is termed recurrence.[346] Using continuous *in vitro* evolution, parameters that could affect recurrence were deduced after comparing many replicate experimental evolutionary lineages with populations of catalytic RNA. For instance, a 150-nucleotide ligase ribozyme was used in one study to predict the likelihood of recurrence happening.[346] Repetitive jumps from one peak in a rugged adaptive landscape to another local optimum were observed on mutations within multiple replicates of actual evolution experiments.[346]

3.1.2. RNA-Cleaving DNA Enzymes. The cleavage of an RNA phosphodiester bond is a relatively facile reaction, which makes it easier to be used than RNA ligation for discovering new reactivities during evolution. Additionally, in comparison to RNA ligation, fewer nucleotides are needed to specify a

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

motif that catalyzes RNA cleavage, thus enabling RNA-cleaving DNA enzymes to be obtained from a small starting population of variants by *in vitro* selection.[315,347] DNA enzymes with relatively high catalytic efficiency, high substrate sequence specificity, and promiscuity toward substrates are critical to boost the catalytic activity of RNA-cleaving reactions, and they have been used in a variety of applications in biochemical and pharmaceutical fields.[315,347]

Using *in vitro* selection, the first metal-dependent DNA enzyme was identified in 1994.[348] First, a starting population of 10^{14} DNAs with each molecule containing a 50 random nucleotide sequence flanked by primer binding sites was constructed. After five successive rounds of selective amplification, individuals that best promote the Pb^{2+}-dependent cleavage of a target ribonucleoside 3′-phosphodiester bond embedded within an otherwise all-DNA substrate were generated with a rate constant of 0.2 min^{-1}.[348] As a result, a simplified version of the catalytic domain that operates in an intermolecular context with a turnover number of 1 min^{-1} was constructed based on the 20 individuals isolated from the final selected population. This activity is about 10^5-fold higher compared to that of the uncatalyzed reaction.[348] Similarly, the approach used to obtain the Pb^{2+}-dependent DNA enzyme was used to obtain a Mg^{2+}-dependent DNA enzyme as well.[349] Substantial catalytic rate enhancements for difficult chemical reactions were achieved. In another practical application, catalytically active DNA enzymes have been used as a unique class of biosensors for Pb^{2+} with a quantifiable detection range from 10 nM to 4 μM and a selectivity of >80-fold over uncatalyzed reaction.[350] In a follow-up study, a colorimetric assay was developed for Pb^{2+} biosensors using DNA enzyme-directed assembly of gold nanoparticles.[351] A simple blue-to-red color change could be visualized directly after the aggregates were disrupted by DNA-catalyzed cleavage. This colorimetric-based biosensor has a detection level ranging from 100 nM to over 200 μM and can be applied to analytes that are subject to *in vitro* selection, thus significantly expanding the scope of nanomaterial applications.[351] The concept of RNA-cleaving DNA enzymes has also been implemented in whole animals. In one study, a DNA enzyme was directed to target the vascular endothelial growth factor receptor 2 mRNA, which mediates angiogenesis associated with tumor growth and metastasis.[352] The enzyme was injected into mice that had been implanted with human breast carcinoma cells. After 4 injections, a nearly 75% reduction in tumor size in the DNA enzyme-treated mice was achieved compared with that of the untreated controls.[352]

3.1.3. Aptamers and Riboswitches. Aptamers are short single-stranded nucleic acids that bind to target molecules with high specificity and affinity.[6,272] The increasing number of aptamers is accompanied by an expanding range of applications such as conditional gene regulation, visualizing RNA and protein distribution, biosensors, and therapeutic agents.[353] Here, we will take the application of gene regulation as an example to describe how to evolve aptamers and riboswitches for improved properties and function.

Aptamers and their natural counterparts, riboswitches, have been widely used as gene expression regulators.[353,354] A regulatory system based on a direct small molecule-RNA interaction with the inducer tetracycline was developed for tetracycline-dependent control of translation in *S. cerevisiae*.[355] The SELEX-selected tetracycline binding aptamers were inserted into the 5′-UTR of a GFP encoding mRNA, resulting in the most efficient (~6-fold) GFP repression near the start codon in the presence of tetracycline. Increasing the thermodynamic stability of the aptamer improved regulation but reduced expression of GFP, while decreasing the stability led to the opposite effect. Sequence variations also influenced the regulation properties of the aptamer; for example, the nucleotide at position 62 affected expression in the absence of tetracycline, the one at position 75 affected regulation in the presence of tetracycline, and the nucleotide at position 21 affected both. Such RNA aptamer-tetracycline ligand-based regulatory systems can be useful tools for conditionally controlling gene expression.

To expand the limited set of existing riboswitches, a riboswitch discovery platform, coupling dual genetic selection and FACS, was developed to create riboswitches that recognize theoretically any ligand of interest.[356] Specifically, the aptamer domain of the ThiM#2 riboswitch was replaced by random 40 nucleotide sequences to construct a library which was then introduced into *E. coli* for multiple rounds of dual selection to find riboswitches responsive to theophylline. As a result, theophylline-responsive riboswitches were identified after only three rounds of selection, and the best riboswitch (Hit 3−5) displayed a 2.3-fold activation of downstream gene expression in the presence of theophylline. Additional random mutagenesis generated improved riboswitches displaying nearly 3-fold activation. Recently, based on growth coupled screening by combining both positive and negative selection, another *in vivo* riboswitch evolution method was established to change the threshold, sensitivity, and dynamic range of riboswitches.[357] Using this method, an *N*-acetylneuraminic acid (NeuAc) riboswitch was successfully evolved to exhibit a higher threshold and a larger dynamic range. The evolved NeuAc riboswitch was then applied to optimize ribosome binding sites and key genes in the NeuAc biosynthetic pathway, resulting in the highest NeuAc production ever reported (14.32 g/L).

Many other aptamers and riboswitches have also been evolved for various ligands for specific applications. Representative examples include guanine recognizing riboswitches,[358] the thiamine pyrophosphate (TPP) riboswitch,[359] a paromomycin-binding synthetic riboswitch,[360] fluorophore-binding RNA aptamers,[361−363] and a pH-responsive riboswitch.[364,365]

3.2. Enzymes

3.2.1. Directed Evolution of Naturally Occurring Enzymes for Chemical Synthesis. Directed evolution has been extensively used to engineer enzymes for improving their catalytic activities and fine-tuning their chemo-, stereo-, and enantioselectivities.[55,366−368] By using directed evolution, numerous enzymes have been engineered for the production of biofuels, materials, fine chemicals, and active pharmaceutical intermediates in various industries.[369−372] Early applications of directed evolution were focused on improving enzyme tolerance to organic solvents, which is necessary to improve substrate solubility for large scale manufacturing.[373,374] In one seminal study, by combining several rounds of random mutagenesis and recombination, Arnold and co-workers created *p*-nitrobenzyl esterase variants with up to 60-fold increase in total activity for hydrolysis of *p*-nitrobenzyl esters in solvents with 30% dimethylformamide.[374] Thermostability is another important property for enzyme engineering. However, it is not unusual that improving thermal stability compromises

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

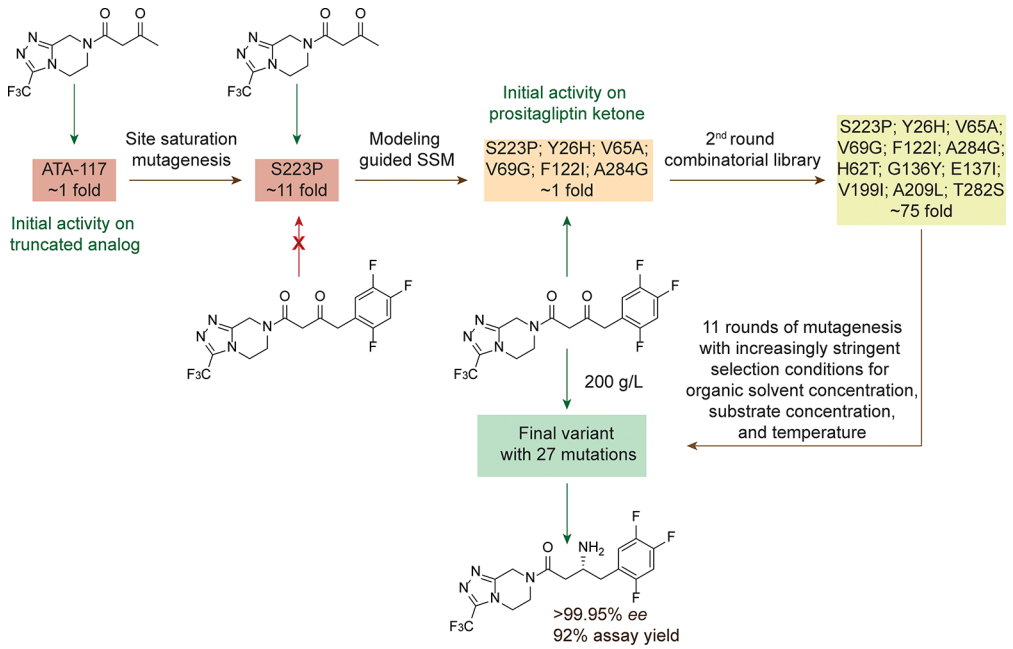
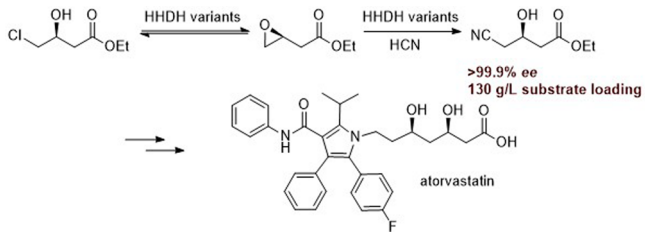


Figure 17. Directed evolution of (*R*)-selective transaminase ATA-117 for asymmetric synthesis of chiral amines from ketones applied to Sitagliptin manufacture.[375]

enzymatic activity at lower temperatures. By carefully controlling the evolutionary path, a thermostable esterase variant was created without the cost of its catalytic activity at lower temperature.[63] One impressive achievement is the development of an *ω*-transaminase for sitagliptin phosphate manufacturing. Choosing a natural transaminase catalyzing *R*-specific transamination of methyl ketones and small cyclic ketones as a starting point, scientists from Merck and Codexis used computational modeling and an *in vitro* coevolution[665] approach to generate a transaminase variant with weak activity that prositagliptin ketone. Multiple rounds of random mutagenesis and screening led to variants with about 25,000-fold improvement in catalytic activity and can produce sitagliptin with >99.95% *ee* at 200 g/L scale, which outcompetes a rhodium-catalyzed process (Figure 17).[375]

In addition to improving the stability, selectivity, and catalytic activity of enzymes, directed evolution has been widely used to engineer enzymes with new reactivity by taking advantage of enzymatic promiscuity. The number of reactions catalyzed by enzymes is far less than that catalyzed by chemical catalysts, which has limited the broad application of enzymes in the chemical industries. Enzyme promiscuity is the ability of an enzyme to catalyze, in addition to its primary reaction, side reactions with different substrates or through different mechanisms. The promiscuous catalytic activities are generally slow compared with the main activity, but they are vital to evolve new enzymes under natural or artificial selection.[11] Enzymes can have promiscuity for reactions that do not exist in nature. Mechanism-guided directed evolution is a powerful strategy to generate new enzymes by exploring protein sequence space and accumulating mutations that enhance the side reaction under certain circumstances.[16] For example, wild-type halohydrin dehalogenase (HHDH) catalyzes the dehalogenation of various aliphatic and aromatic vicinal halo alcohols.[376] A mechanistic study of HHDH revealed that the active site can accommodate an epoxide and other small negatively charged ions such as an azide instead of a halide. The catalytic tyrosine residue can work as a catalytic acid for epoxide protonation when the reactions are performed under acidic conditions. The changes of substrate and reaction condition can trigger the reverse reaction, epoxide-ring opening.[377] In 2007, scientists at Codexis performed directed evolution of the HHDH from *Agrobacterium radiobacter*, resulting in variants used for large-scale synthesis of ethyl (*R*)-4-cyano-3-hydroxybutyrate, a building block of atorvastatin, with >99.9% *ee* at a substrate loading of 130 g/L. The improvement in productivity was approximately 4,000-fold (Scheme 1).[180]

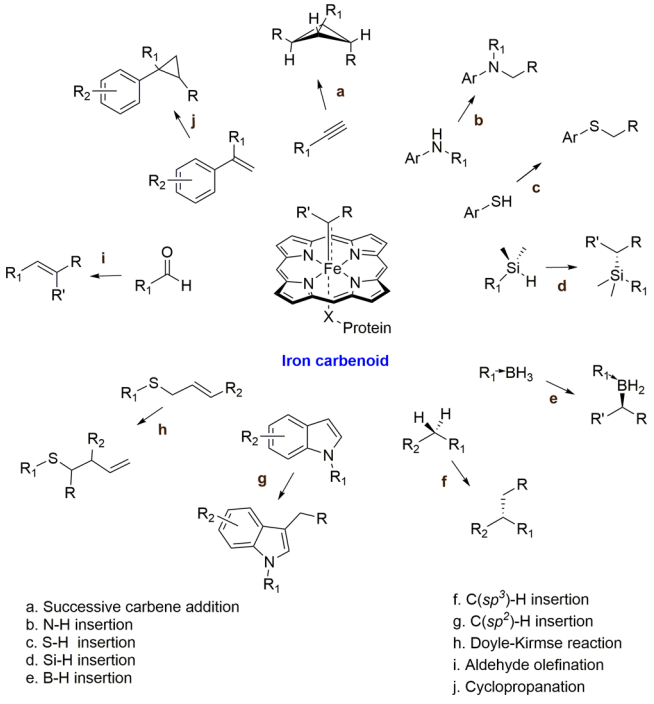
Scheme 1. Application of Halohydrin Dehalogenase HHDH in the Synthesis of Atorvastatin



Directed evolution is a powerful tool to uncover metalloenzymes' promiscuous activity to catalyze those reactions that were only catalyzed by transition metal catalysts before. Transition-metal catalyzed C═C and C−H functionalizations by carbenoid transfer are widely used to synthesize natural products and pharmaceutical intermediates.[378,379] Inspired by the well-established capacity of metalloporphyrin complexes for cyclopropanation reactions, Arnold and co-workers engineered $P450_{BM3}$ variants that can catalyze highly diastereo- and enantioselective cyclopropanation and N−H insertion by carbene transfer (Scheme 2).[380−384] More recently, cytochrome c from *Rhodothermus marinus* was identified with catalytic promiscuity for Si−H insertion of phenyldimethylsilane with very low total turnover number (TTN) but good

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Scheme 2. Carbene Transfer Reaction Catalyzed by Evolved Hemoproteins

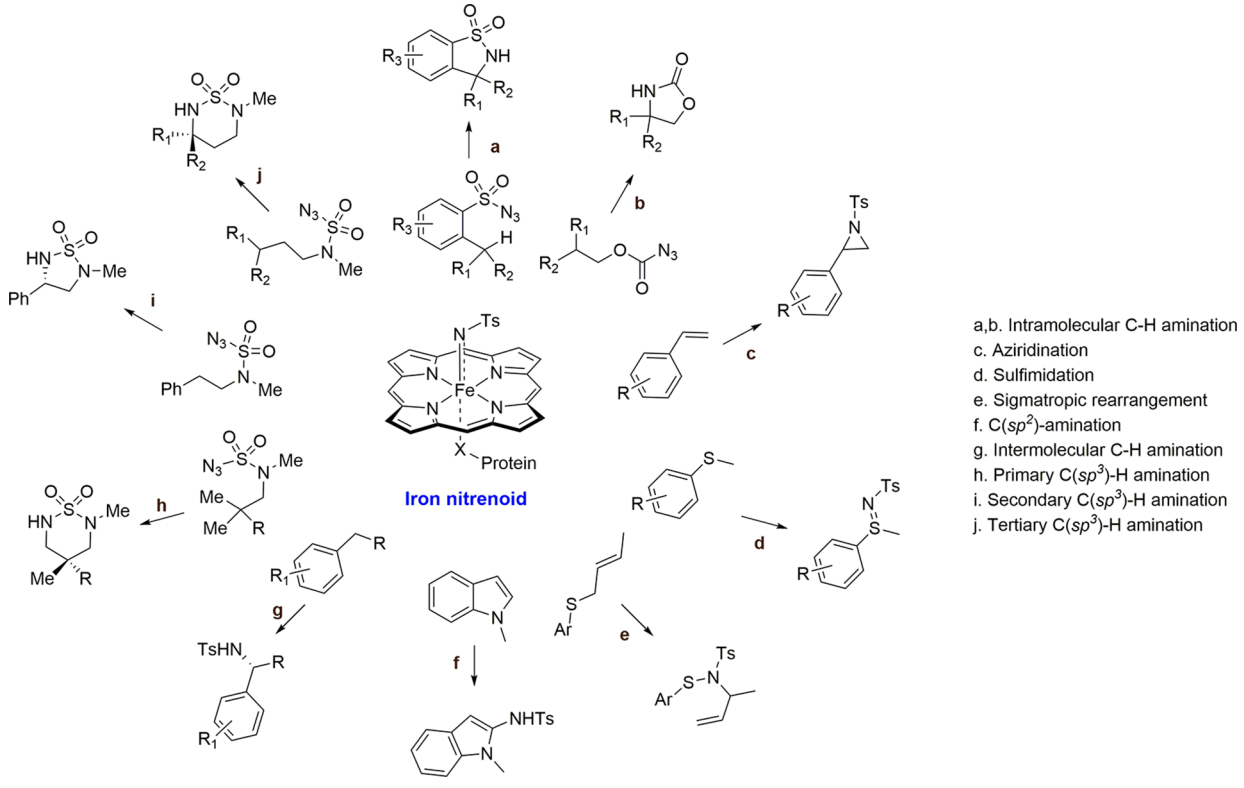


enantioselectivity. Directed evolution focusing on the active site residues resulted in a triple variant with 15-fold improved TTN for the synthesis of carbon−silicon bonds in a broad range of substrates with up to 99% *ee*.[385] Other residues at the same sites catalyzed boron−silicon bond formation through carbenoid B−H insertion. Various organoborons have been prepared on a gram-scale (up to 15,000 TTN) with excellent ee using whole-cell catalysts containing different *Rma* cyt c variants.[386,387] Those studies have expanded the scope of boron and silicon chemistry in living systems since enzymatic reactions for Si−C and B−C bond formation have not been identified in biology so far (Scheme 2).[387] Except for P450s, the evolved oxygen-transport protein myoglobin can also catalyze selective carbene transfers such as cyclopropanation,[388] N−H insertion,[389] S−H insertion,[666] benzofuran cyclopropanation,[390] and intramolecular cyclopropanation to construct cyclopropyl-*δ*-lactones (Scheme 2).[391]

Directed evolution can expand the ability of enzymes to construct C−C bonds, which is challenging but essential for improving their applications in chemistry and synthetic biology. sp^3 alkylation is normally catalyzed by expensive transition-metal catalysts.[392−394] Iron-catalyzed reactions require very high temperature, accounting for the high activation energy barrier for carbene insertion into a C−H bond.[395] Impressively, evolved P411 (P450 with the axial cysteine replaced by serine) variants can catalyze abiological sp^3 C−H functionalization on substrates containing benzylic, allylic, or *α*-amino acid with high turnover number and excellent selectivity at room temperature (Scheme 2). This study demonstrated that the protein framework can confer the activity to an unreactive cofactor.[396] Engineered P411 variants were used to catalyze C−H functionalization for enantiodivergent C(sp^3)−H fluoroalkylation to access organofluorine compounds that are currently not accessible with small-molecule catalysts.[397] C(sp^2)−functionalization of unprotected indoles was also achieved by both engineered Mb and P411 variants (Scheme 2).[398,399]

Furthermore, directed evolution can be used to create enzymes with higher catalytic activities than the transition

Scheme 3. Nitrene Transfer Reactions Catalyzed by Evolved Hemoproteins



https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

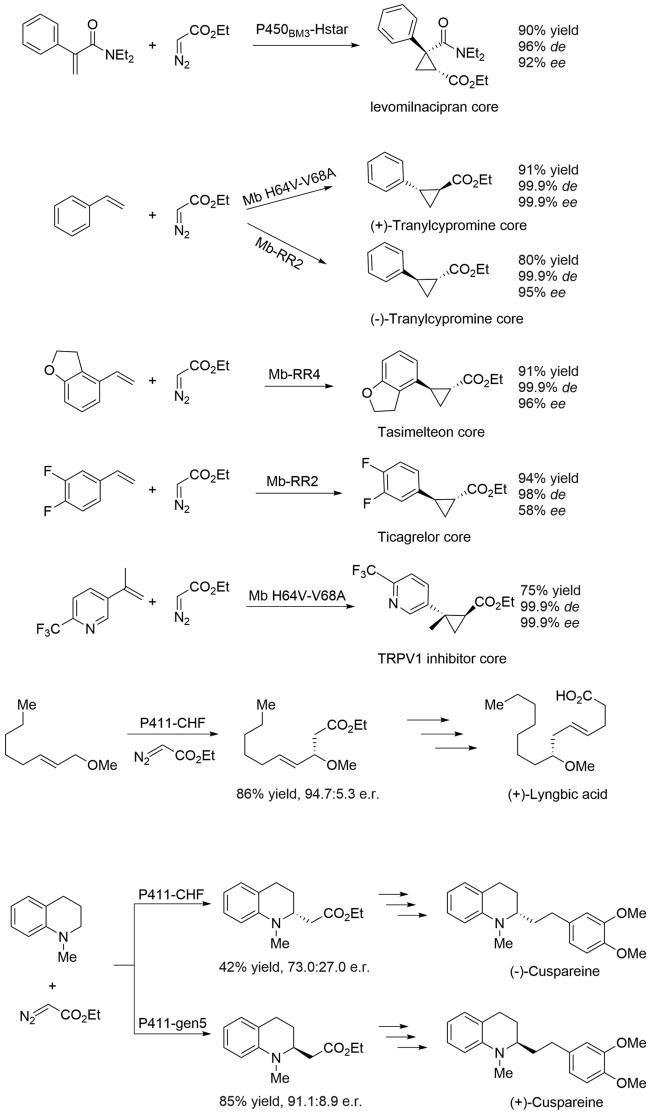
metal catalysts for the same reactions. In addition to carbene transfer reactions, transition-metal-catalyzed nitrene transfer is another crucial strategy to synthesize C−N bonds in many bioactive synthetic and natural compounds. Iron-containing enzymes were also engineered to catalyze nitrene transfer reactions, such as regioselective intramolecular C−H amination on sulfonylazide compounds,[400−402] sulfimidation of sulfides,[403] sigmatropic rearrangement for enantioenriched allylic amines synthesis,[404] and enantioselective olefin aziridination (Scheme 3).[405] Interestingly, the evolution for sulfimidation yielded P411 variants that also catalyzed intermolecular C−H amination of alkanes.[406] Further site-saturation mutagenesis around the active site resulted in $P411_{CHA}$ with up to 1,300 TON for enantioselective C(sp^3)−H amination (up 99.9% *ee*), exceeding any chiral transition metal catalyst reported so far.[406] The synthetic power of P411 variants was further expanded recently by catalyzing chemo- and enantioselective primary C(sp^3)−H amination with hydroxylamine esters as nitrogen sources for a broad range of substrates at a preparative scale, which has never been discovered in biological systems.[407]

Directed evolution can also be used to engineer enzymes to catalyze abiotic reactions that are challenging for chemical catalysts. For example, ring strained structures such as cyclopropenes and bicyclobutane are particularly attractive intermediates in chemical and materials synthesis. However, their synthesis is demanding, requiring expensive chiral transition metal catalysts.[408,409] Asymmetric bicyclobutanes with multiple chiral centers are a good example.[410] *E. coli* harboring engineered P411 variants can be used to prepare a broad range of cyclopropenes with different functional groups with good enantioselectivity. Both enantiomers can be obtained by using variants from different evolutionary trajectories. Remarkably, variants were also identified that catalyze successive carbene addition to unsaturated carbon−carbon bonds to form a wide range of bicyclobutanes on a preparative scale (Scheme 2).[411,412] In another example, using *E. coli* harboring truncated $P411_{Diane1}$ variants as the whole-cell catalysts, Arnold and co-workers reported chemoselective asymmetric amination of primary, secondary, and tertiary C(sp^3)−H bonds. The enzymes enable the enantioconvergent transformation of racemic tertiary C(sp^3)−H bond to generate tetrasubstituted stereocenters that are also inaccessible with small molecule catalysts (Scheme 3).[413]

Overall, engineered hemoproteins have achieved several important chemical transformations widely used in synthetic chemistry but cannot be found in nature, thereby increasing enzymes' potential for industrial applications. Several engineered hemoproteins have been demonstrated to synthesize high-value compounds in a preparative scale (Scheme 4). First, a highly active His-ligated variant of $P450_{BM3}$-Hstar was used to synthesize the core of levomilnacipran in one step at high isolated yield and *ee*.[414] Additionally, the core structures of cyclopropane-containing drugs such as Tranylcypromine, Tasimelteon, Ticagrelor, and a TRPV1 inhibitor were also prepared on a gram-scale with excellent diastereo- and enantioselectivity by using *E. coli* harboring myoglobin variants.[371,372] P411 variants catalyzing sp^3 functionalization were used to prepare a chiral precursor of the valuable natural product lyngbic acid and both enantiomers of cusparieine.[396]

3.2.2. Directed Evolution of Artificial Enzymes for Chemical Synthesis. *3.2.2.1. De Novo Artificial Enzymes.* Advances in computational chemistry and biology have

Scheme 4. Prepscale Synthesis of High Value Compounds by Engineered Hemoproteins

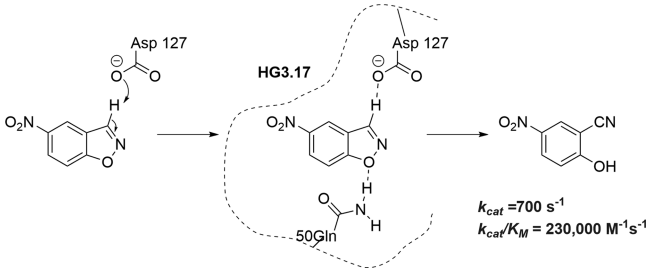


enabled the *de novo* design of enzymes that catalyze abiotic reactions.[415] The "inside-out" approach, by which the active site designed for a target reaction is incorporated into a stable protein scaffold, has been successfully demonstrated to produce catalytically active and selective proteins.[416] The resultant *de novo* enzymes are stable, but their kinetic performance generally falls short of that of the natural equivalents. Subsequent structure- and mechanism-guided site-directed mutagenesis and random mutagenesis can significantly improve the enzyme catalytic efficiency and turnover number to approach the natural counterparts.[417] For example, Kemp eliminases were designed to catalyze base-promoted deprotonation of 5-nitrobenzisoxazole, which serves as a model reaction for understanding the deprotonation of carbon—a process that has considerable kinetic and thermodynamic barriers but frequently occurs in a living organism.[418] One of the variants from HG3, a structurally stable *de novo* Kemp eliminase with a k_{cat}/K_m around 430 M^{-1} s^{-1} was randomly mutagenized to identify "hot spots". Focused mutagenesis on both the "hot spots" and the substrate-binding sites led to a variant with k_{cat} = 700 s^{-1} and k_{cat}/K_m = 230,000

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

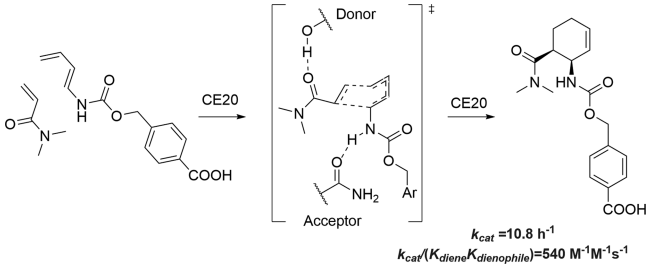
M^{-1} s^{-1}, approaching the exceptional catalytic efficiency of highly optimized natural enzymes such as triosephosphate isomerase (Scheme 5).[419]

Scheme 5. Kemp Elimination



Although computational simulation and modeling can optimize the performance of *de novo* enzymes to a certain extent, directed evolution is still needed since it provides complementary solutions for enzyme optimization. A Diels–Alderase was created by positioning a hydrogen bond acceptor and donor in the apolar binding pocket of a rigid β-propeller scaffold that is highly evolvable.[420] Installation of a helix-turn-helix lid element through computational refinement or using directed evolution can both improve the Diels–Alderase's catalytic efficiency but in different ways.[421,422] A lid element mainly increases the substrate's binding affinity but keeps the turnover number unchanged while directed evolution improves both. Combining mutations from computational prediction and directed evolution resulted in a variant with catalytic efficiency between its precursors because of the compensation effects of turnover and binding affinity. However, it served as a new starting point for evolution, which resulted in a new variant that was several times more efficient than both precursors. This new variant turned out to be the most proficient Diels–Alderase at that time, which can selectively synthesize (3*R*,4*S*)-*endo*-cyclohexene from 4-carboxybenzyl-*trans*-1,3-butadiene-1-carbamate and *N*,*N*-dimethylacrylamide at a preparative scale (Scheme 6).[422] Most recently, Hilvert

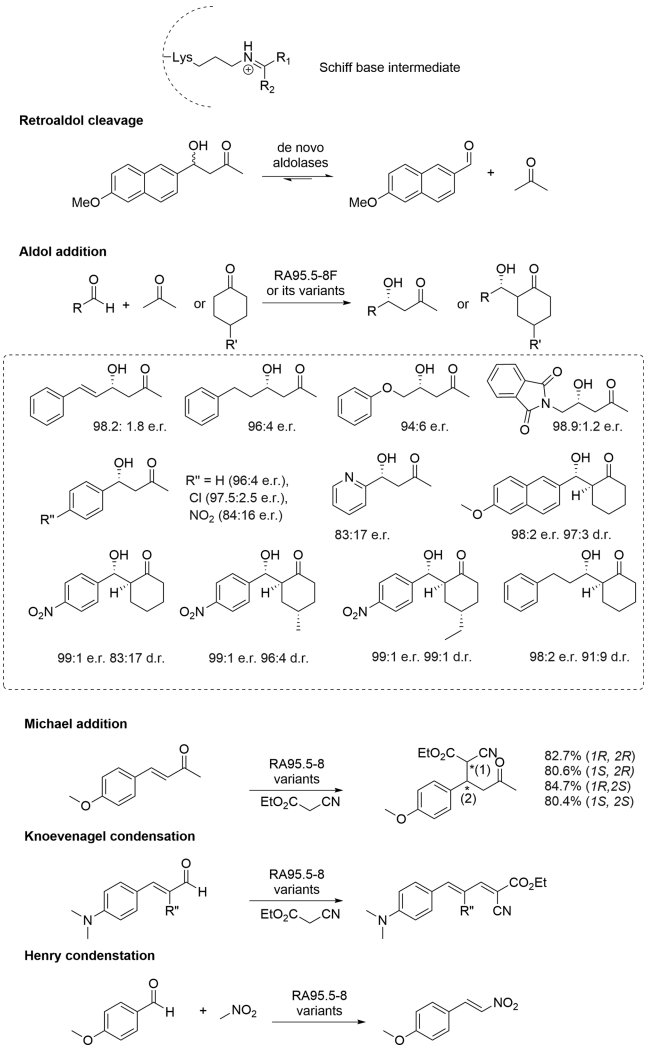
Scheme 6. Diels–Alder Reaction



and co-workers evolved a *de novo* zinc-containing peptide into a highly active catalyst for a biological hetero-Diels–Alder reaction between azachalcone and 3-vinylindole with a catalytic proficiency up to 2.9 × 10^{10} M^{-1} that is much higher than the artificial metalloenzyme created by using natural protein scaffolds.[423] This study showed that unlike natural proteins that have been optimized by natural evolution for specific functionality, *de novo* protein scaffolds with promiscuous binding pockets for metal cofactors and substrates might be more amenable to function diversification.

The general principles for directed evolution can also be applied to *de novo* enzymes. For example, an increase in the library size and throughput of mutagenesis results in a higher possibility of finding optimum variants. Divergent evolution can uncover *de novo* enzymes' promiscuous activity that is not from the design intention. Retro-aldolases are probably the most mechanistically complex *de novo* enzymes developed so-far. They catalyze the cleavage of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone to 6-methoxyl-2-naphthaldenyde and acetone through a multistep pathway involving amine catalysis and enzyme-bound Schiff base intermediates (Scheme 7).[424]

Scheme 7. Reaction Catalyzed by Retro-Aldolase through Schiff Base Intermediate



One of the initial aldolases, such as RA95.0, catalyzes the reaction with 15,000-fold rate acceleration over the background. However, its k_{cat}/K_m is only 0.19 M^{-1} s^{-1} and k_{cat} is only 0.1 × $10^{-3}s^{-1}$, which are too low to make it a useful catalyst. Several rounds of directed evolution using both random and focused mutagenesis coupled with a microtiter plate-based screening method boosted its activity by more than 4,400-fold, affording catalytic efficiency approaching those of natural enzymes.[425] Further evolution with a ultrahigh-throughput droplet-based microfluidic screening platform led to a new variant, RA95.5-8F, with an additional 30-fold improvement in catalytic efficiency, rivaling the efficiency of

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384–12444

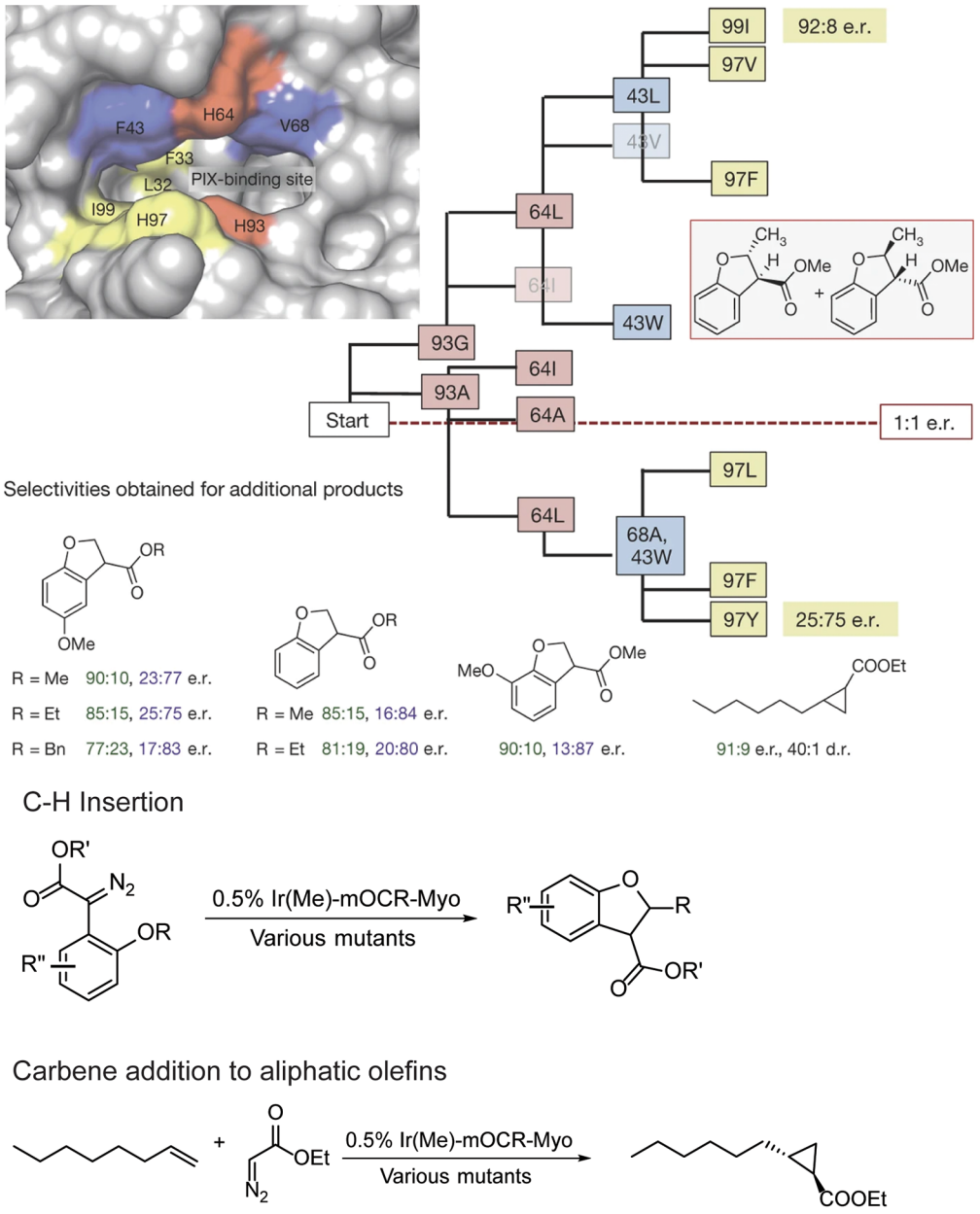


Figure 18. Directed evolution strategy used to obtain Ir(Me)-mOCR-Myo variants capable of producing either enantiomer of the products of C−H insertion reactions of varied substrates.[438] (Reproduced with permission from ref 438. Copyright 2016, Springer Nature. Original Figure 4 was reproduced in this figure.)

class I aldolases (Scheme 7).[426,427] Notably, this new variant can catalyze stereoselective C−C bond formation by reversible aldol reaction over a broad substrate scope.[427,428] Divergent evolution of this *de novo* aldolase resulted in variants that could catalyze promiscuous C−C bond formation reaction promoted by organocatalysts such as Knoevenagel condensation,[429] Michael addition,[430] and Henry condensation.[431] These promiscuous variants can serve as the starting point for further evolution to improve enzyme activity and selectivity.[429,432]

3.2.2.2. Artificial Metalloenzymes. A wide variety of metal complexes in organic and organometallic chemistry have been used to catalyze reactions that natural enzymes cannot perform. Unlike enzymes, these catalysts tend to have limited enantio-, regio-, and chemoselectivity and require harsh conditions. Creating artificial metalloenzymes (ArMs) by either replacing the metalloenzymes' native cofactors with the artificial ones or anchoring metal complexes in native scaffolds is an effective strategy to expand the activities of native enzymes and improve the selectivity of chemocatalysis reactions.[433] However, the activity and selectivity of ArMs remain suboptimal and need to be further improved by directed evolution. Using native proteins as scaffolds to design ArMs is an effective strategy because the native proteins are highly evolved and have relatively stable structures that are amenable to physiological condition changes and further directed evolution.

Approximately half of the characterized proteins contain a metal cofactor as part of their catalytic function.[434] A native metalloenzyme can serve as a great scaffold for designing ArMs using abiotic cofactors because it already has the binding site for metal ions or metal complexes. Heme-containing proteins are widely distributed in living organisms and have been

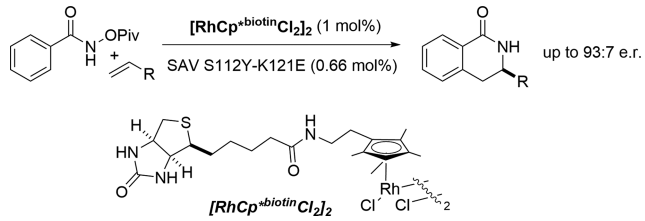
https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

engineered for diverse chemical transformations. Since the number of chemical reactions catalyzed by free metal−porphyrin complexes of Ru, Rh, and Ir are much greater than that of the free Fe analogues, replacing Fe with Ru, Rh, or Ir in hemoproteins can result in new chemistry.[435−437] Artificial myoglobin was constructed by expressing apo-myoglobin in *E. coli* under proper conditions, followed by reconstitution with metal−porphyrin IX ([M]-PIX) cofactors with M being Fe, Co, Cu, Mn, Rh, Ir, Ru, and Ag. Among them, Ir(Me)-mOCR-Myo was identified as the most active in functionalization of C−H bonds to form C−C bonds by carbene insertion and cyclopropanation. Focused mutagenesis was used to generate Ir(Me)-mOCR-Myo variants that generate either enantiomer of C−H insertion products and also catalyze enantio- and diastereoselective cyclopropanation on β-substituted vinylarenes and unactivated aliphatic α-olefins, which cannot be realized by any natural enzymes at that time (Figure 18).[438] A similar strategy was used to prepare Ir(Me)-CYP119 and site-saturated mutagenesis was used to engineer its substrate binding site. The resultant quadruple variant Ir(Me)-CYP119-max (C317G, T213G, L69V, V254L) can catalyze carbene insertion of C−H with a catalytic efficiency comparable to that of CYP119 on native substrates.[439] Additional directed evolution further expanded the substrate scope of C−H functionalization catalyzed by Ir(Me)-CYP119 variants over additional terminal and internal, activated and inactivated, electron-rich and electron-deficient, conjugated and nonconjugated alkenes and phthalan derivatives with the high selectivity that is hard to achieve by small-molecule catalysts.[440,441]

Not all metal ions or complexes can readily replace the native cofactors of metalloenzymes. Covalently anchoring is an alternative to link unfit metal complexes to any scaffold protein, including those without native cofactors. The resulting ArMs normally have very low or even no activity due to the absence of the entrance point or defined binding site of the substrates.[442,443] SSM and random mutagenesis are therefore essential tools to develop efficient ArMs from unreactive proteins as ArM scaffolds. Dirhodium complexes can catalyze a wide range of reactions including cyclopropanetrione and X-H insertion (X = C, N, O, S, Si). However, the low selectivity remains as a challenge for many reactions.[444] To resolve this issue, a dirhodium ArM was created by linking tetramethyl m-benzenedipropionic acid ligands-based dirhodium cofactor to a prolyl oligopeptidase scaffold (POP) with an L-4-azidophenylalanine residue by strain-promoted azide−alkyne cycloaddition (SPAAC) reaction.[445] Since POP was selected based on its thermal stability and large hydrophobic pocket rather than metal complex binding affinity, variant POP-ZA_4 was first generated to open the entrance for cofactor and substrates. SSM was used to modify the second coordination sphere of the cofactor, improving its substrate specificity and enantioselectivity of styrene cyclopropanation.[446] Additionally, a random mutagenesis platform was developed based on 96-well plates for ArMs library preparation and screening.[447] From a library of variants created by epPCR and combinatorial codon mutagenesis, new variants were identified with beneficial mutations far from the metal binding sites, which further improved the enantioselectivity on olefin cyclopropanation. New variants catalyzing N−H, S−H, and Si−H carbene insertion can also serve as starting points for further directed evolution.

Biotin−avidin technology has been widely used to non-covalently incorporate a metal catalyst within a protein of interest through the strong interaction between biotin and (strept)avidin (M-biotin-SAVs) ($K_a \sim 1 \times 10^{14}$ M^{-1}).[448] The binding affinity is not altered by using biotin with different valeric acid side chains, which ensures the integrity of the diverse organometallic complexes (Rh, Ru, Pd, Ir, Os, V) and allows orthogonal generation to optimize ArMs via both SAV evolution and organometallic ligand diversification.[449,450] Relying on this strategy, multiple ArMs have been created for hydrogenation,[450] transfer hydrogenation,[451−454] allylic alkylation,[455] dihydroxylation,[456] sulfoxidation,[457] alcohol oxidation,[458] Suzuki cross-coupling,[459] C−H activation,[460] and olefin metathesis.[448,461] SSM was applied to S112, K121, and/or L124 of (strept)avidin to create enantioselective variants. For example, M-biotin-SAVs-catalyzed C−C bond formation reactions are invaluable since there is no equivalent in the natural enzymatic repertoire. Rh(III)-biotinylated SAV-S112Y K121E was found to be the best performer in the selective coupling of benzamides and alkenes to dihydroisoquinolone, achieving a 100-fold faster rate than that of isolated Rh complexes as well as enantiomeric ratios (er) up to 93:7 (Scheme 8).[460]

Scheme 8. C−H Activation



To facilitate the directed evolution of M-biotin-SAVs, a streamlined protocol was established for parallel engineering of dozens of transfer hydrogenation reactions or metathesis reactions.[462] After the protein library was expressed in *E. coli* in 96 deep well-plates, the cell lysates were treated with diamide, a glutathione scavenger,[463] followed by adding metal cofactors and subsequently performing the transfer hydrogenation reactions. Using this platform, nine amino acid residues within 10 Å of the iridium center in [Cp*Ir(biot-p-L)Cl]⊂SAV K121A were chosen for iterative saturation mutagenesis, which resulted in variants for both *S* and *R* products in high yield.[464] For metathesis, SAV was immobilized on Sepharose iminobiotin beads under basic conditions and cell debris was removed by centrifuge. The treatment of beads under acid released SAVs and M-biotin-SAVs were constructed by adding metal cofactors in the reaction buffer. This system is versatile and could be adapted to evolve other M-biotin-SAVs.[464] Notably, an *in vivo* directed evolution platform was established recently, which greatly improved the throughput of M-biotin-SAVs engineering (Figure 19).[465] The SAV variants were expressed in *E. coli* and secreted into the cell's periplasm through fusion to an N-terminal signal peptide from the outer membrane protein A, thus eliminating inhibitor scavenging effects. *In vivo* reconstruction of Ru(II)-biotin-SAV variants was performed by incubating the cells with metal cofactors followed by washing. Bio-orthogonal metathesis can be performed using the whole cells directly. As more examples of abiotic reactions catalyzed by M-biotin-SAVs have been demonstrated in *vivo*,[466] *in vivo* directed evolution to create

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

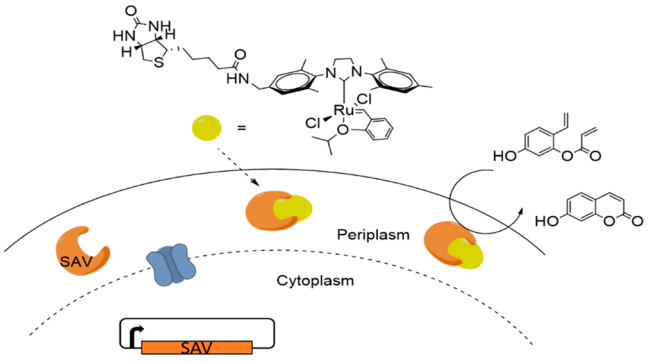


Figure 19. Streptavidin-based artificial metalloenzymes for *in vivo* metathesis and directed evolution.[465] (Adapted with permission from ref 465. Copyright 2016, Springer Nature. Original Figure 1a was adapted in this figure.)

new catalysts with new reactivity or improved activity might become possible.

3.2.2.3. Enzymes with Noncanonical Amino Acids. The structural and functional diversity of the 20 canonical (cAAs) amino acids of enzymes is relatively small compared with diverse chemical catalysts for organic synthesis. However, with the maturing methods of reprogramming the genetic code to incorporate noncanonical amino acids (ncAAs) into proteins, more and more ncAAs have been used to improve the catalytic activity and selectivity of enzymes.[467] Directed evolution has also been adapted to an expanded genetic code for improving catalytic efficiency of enzymes with ncAAs.[468,469] For example, Green and co-worker generated an artificial hydrolytic enzyme that used N_δ-methylhistidine as a ncAA nucleophile to catalyze the hydrolysis of fluorescein, which cannot be achieved by using artificial enzymes with canonical nucleophiles.[470] Several rounds of directed evolution were applied to improve its hydrolysis efficiency by 9,000-fold higher than free N_δ-methylhistidine in solution and 2,800-fold higher than common organocatalysts dimethylaminopyridine and *N*-methylimidazole. Instead of evolving cAAs around ncAAs catalytic center, Schultz and co-workers developed a general strategy to randomly substitute any cAAs of β-lactamase with a set of substituted aromatic ncAAs and identify a unique *p*-acrylamido-phenylalanine mutant with improved catalytic efficiency that cannot be reproduced by any of the 20 cAAs at the same location. This study suggested that expanding the building blocks of proteins may offer unique solutions to organisms in the evolution of new functions.[471]

3.2.2.4. Programmable Nucleases for Gene Editing. Precise gene editing techniques play a key role in molecular medicine and synthetic biology. The development of programmable nucleases such as zinc-finger nucleases (ZFNs), TALENs, CRISPR, and the CRISPR-associated nuclease Cas gave rise to the genome engineering field and brought gene editing from conceptual studies to clinical trials.[472] Both ZFNs and TALENs are artificial nucleases, created by linking a specific DNA-binding domain (DBD) to the nonspecific bacterial endonuclease *Fok*I. Zinc fingers (ZFs) are DBDs approximately 30 amino acids long and recognize three nucleotides in the major groove of DNA.[473,474] Several strategies have been developed to produce ZFs with tailor-made DNA specificities.[475,476] A highly efficient *Fok*I cleavage domain was evolved with improved nuclease activity.[477] Transcription activator-like effectors (TALEs) are more programmable DBDs than ZFs and consist of an N-terminal domain followed by a series of 33−35 amino acid long tandem repeats, a nuclear localization sequence, a transcription activation domain, and a C-terminal domain.[478,479] Amino acids at positions 12 and 13 of each repeat recognize and bind to a specific DNA base.[480,481] Off-target effects in a genome and constraints on the recognition site starting with T are two major limitations of TALEs. Liu and co-workers used PACE to evolve TALEs with improved DNA cleavage specificity and improved activity at 5′ A and 5′C sites relative to canonical TALE proteins.[324] Importantly, as a general continuous *in vivo* evolution platform, PACE can be used to engineer the specificity of other DBDs.

The CRISPR-Cas system is one of the most important technologies for gene editing in a wide variety of organisms and cell types. Among various CRISPR-Cas systems, Cas9 from *Streptococcus pyogenes* is one of the most widely used systems for targeted gene disruption, transcriptional activation and repression, epigenetic modification, and base editing in a broad range of organisms. Unlike ZFNs and TALENs, SpCas9 has DNA recognition and nuclease machinery in a single protein. It recognizes the target site by a programmed chimeric sgRNA that encodes a sequence complementary to a target protospacer. Additionally, the target site must have a NGG PAM sequence to support Cas9 recognition.[123] Based on the crystal structure and gene editing mechanism, several Cas9 variants have been created with enhanced specificity and altered PAM sequences.[482−485] However, some variants show improved specificity at the cost of reduced on-target activity.[486−489] Recently, Sniper-Cas9 and evoCas9 were created by directed evolution in *E. coli* and yeast respectively and showed improved specificity in human cells, while maintaining near wild-type on-target editing efficiency.[490,491] Impressively, Liu and co-workers used PACE to obtain xCas9 that recognizes a broad range of PAM sequences including NG, GAA and GAT. Meanwhile, xCas9 is more specific than SpCas9 with much lower off-target activity at both NGG and non-NGG target sites without a trade-off on on-target activity.[325] PACE was also used to evolve cytosine base editors with both APOBEC1 and CDA1 deaminases for more efficient editing of high GC loci while maintaining editing efficiency in other sequence contexts.[492]

3.2.3. Binding Proteins. *3.2.3.1. G Proteins.* G protein-coupled receptors (GPCRs) are the most important family of proteins as targets of therapeutics development.[493] It is estimated that approximately 30% of all FDA approved drugs use GPCRs as targets.[494] GPCRs are attractive for biomedical and drug-discovery research due to their importance in intracellular signaling and relevance to diseases such as cancer, infection, and inflammation.[495] Even though GPCRs are important drug targets, there is relatively little information about their structures because it is difficult to isolate GPCRs due to their instability and toxic effects on host cells; crystallization is also extremely difficult due to the low yields of purified GPCRs.[496] Therefore, many engineering efforts have been dedicated to improve the expression level of GPCRs, including attempts to optimize the parameters influencing protein expression, such as host cells, expression plasmids, and induction conditions.[497] Directed evolution was successfully used to improve the expression of GPCRs.[498−500] For example, Sarkar and co-workers engineered a functionally expressed GPCR in *E. coli* through directed evolution using rat neurotensin receptor-1 (NTR1) as a model system. The

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

engineered variant has improved expression, stability, and binding selectivity. The expression level of a randomized gene library was screened by a radioligand assay, in which the receptors were saturated with a fluorescent agonist. The most fluorescent cells were isolated by FACS. This work presents a high-throughput platform for the directed evolution of GPCRs to improve protein expression level and stability while maintaining protein function and selectivity. Such a method can potentially be applied to other membrane proteins and achieve milligram-level production. For example, three human GPCRs, tachykinin receptor NK_1, α_{1a}AR, and α_{1b}AR, were engineered for higher expression levels and increased stability by directed evolution.[498] More recently, Plückthun and co-workers engineered functionally expressed human oxytocin receptor (OTR) using directed evolution.[501] OTR is clinically targeted due to its biological role in organizing sexual reproduction and social behavior. However, OTR is one of the most challenging GPCRs due to its poor stability and toxicity to *E. coli* cells. In this study, Plückthun and co-workers created a variant library using epPCR and selected variants with improved functional expression first in yeast and then in *E. coli*.

3.2.3.2. Antibodies. Antibodies are Y-shaped immunoglobulin proteins produced by B lymphocytes. They have a defined structure and the ability to bind molecular targets with high affinity and specificity.[502] Since the first monoclonal antibody for therapeutic purposes was approved by the FDA in 1986,[503] there have been a huge research interest in the biotechnological and therapeutic fields for engineering antibodies. By March 2017, 60 therapeutic antibodies had been approved by the FDA.[504] Besides therapeutic applications, antibodies are widely used as research tools. For example, they are used to visualize molecules unseen by naked eye and in routine techniques such as Western blot and flow cytometry. They are also important for in many diagnostic techniques,[505] including the differentiation of cancer and noncancer cells.[506] With the potential to bind almost unlimited targets with desired specificity, many studies were performed to engineer antibodies using directed evolution. One of the major engineering goals is to improve antigen-binding affinity. Wittrup and co-workers successfully engineered a single-chain antibody variant with a two order of magnitude higher ligand-binding affinity than any reported variants at the time and a 1,000-fold decrease in the rate of dissociation through directed evolution. The single-chain scFv library was created by error-prone DNA shuffling, and the resulting library of yeast surface displayed antibodies was screened by flow cytometry sorting.[507] Directed evolution was also employed together with ribosome display and mRNA display.[508−510] Due to their high binding specificity, antibodies can also be used as tools to study other proteins. For example, Kobilka and co-workers engineered Nb80, a conformation-selective single-domain camelid antibody, through directed evolution. The nanobody was then used to stabilize the active-state of β_2-adrenoceptor, which was used to obtain the crystal structures of the activated receptor.[511]

As one of the various antigen targets of antibodies, T cell receptors are an essential component of adaptive immunity. Thus, T cell receptors (TCRs) have attracted intense attention for T-cell-based therapies.[512] One major engineering target of TCRs is stability. Many TCRs have low solubility and tend to aggregate, resulting in low or variable production of the recombinant proteins. Wittrup and co-workers reported a single-chain T cell receptor (scTCR) scaffold with high stability and soluble expression using directed evolution. The TCR library was created by epPCR and screened by yeast surface display. The top variant was stable up to 46 °C, while maintaining specificity.[513] Another engineering property is affinity, since low affinity is a major limitation of TCR-based therapeutic and diagnostic applications. Jakobsen and co-workers reported a high affinity human TCRs specific for two peptide-human leukocyte antigen complexes. They constructed a library of TCRs by introducing random mutations and screened by phage display. After several rounds of directed evolution, they increased the affinity of the engineered TCR by $\sim 10^6$ fold for targeting of the cell-surface peptide human leukocyte antigen (pHLA).[514]

3.2.3.3. Transcription Factors. Transcription factors (TFs) regulate gene expression by either activating or repressing the transcriptional machinery according to different environmental conditions. TFs are widely used as gene regulatory tools in metabolic engineering, gene therapy, and biosensing.[515] However, due to their limited substrate scope, specificity, and sensitivity, native TFs been extensively engineered. Arnold and co-workers used directed evolution to engineer the quorum-sensing transcriptional activator LuxR for enhanced signaling specificity. EpPCR was used to create a library of genes encoding LuxR, which were analyzed with a positive−negative dual selection system. The engineered variant R67 M achieved a 50,000-fold change of specificity compared with wild-type LuxR.[516] In addition, Cirino and co-workers reported a successful case of altering AraC inducer specificity using directed evolution. In the study, two AraC saturation mutagenesis libraries were created, followed by high-throughput screening by FACS. The reported variant showed altered specificity toward D-arabinose and no response toward the native effector L-arabinose. Moreover, the variant was not induced by other sugars such as D-xylose, D-fucose, and D-lyxose.[517] This work demonstrates the potential of designing customized gene switches using directed evolution. Other works also demonstrated the use of directed evolution to modify TFs for metabolic engineering purposes.[518,519]

3.2.4. Transporters. Transporter proteins, located in cell membranes, have a pivotal role in controlling the influx of substrate and efflux of products for small molecule production.[520] Thus, there are many studies of transporters using directed evolution. Cellodextrin transporter 2 (CDT2) is considered to have more potential than CDT1 for biofuel production under anaerobic conditions. However, it is limited by the low activity. Zhao and co-workers engineered CDT2 for higher uptake of cellobiose and improved cellobiose fermentation. The CDT2 library was constructed by epPCR and screened by a fluorescence assay measuring cellobiose uptake. The evolved variants showed ∼4-fold increase in cellobiose consumption rate and ethanol productivity.[521] Recently, Nielsen and co-workers performed directed evolution on the membrane transporter Tpo1 to facilitate cellular production of medium-chain fatty acids (MCFAs). MCFAs are valuable platform chemicals. However, their production by microbial biocatalysts is limited by the cellular toxicity of MCFAs. In this study, a library of Tpo1 variants was generated by epPCR and selected via a growth-coupled assay for increased activity conferring cellular resistance against C10 fatty acids. It was found that the yeast strain expressing the best TPO1 variant increased the titers of C10, C12, and C14 fatty acids by 0.8−2.1-fold compared to the yeast strain expressing

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

the wild type Tpo1.[522] In addition, directed evolution has been proven to be successful in engineering transporters for improved efflux pump.[523,524]

3.2.5. Reporters. GFP is arguably the most widely used reporter protein for gene expression and has many other applications such as fluorescence microscopy.[525] Due to the limited brightness, wavelength, and folding stability, applications of native GFP were limited. One of the pioneering directed evolution studies on GFP was performed by Stemmer and co-workers in 1995. After three rounds of DNA shuffling and a visual screening under UV light, a variant with 45-fold greater brightness than the commercial GFP at that time was successfully obtained.[57] Due to the ease of constructing screening assay of GFP, directed evolution is often employed for engineering GFP for higher intensity[526−529] and sensitivity.[530] Besides green fluorescent protein, other types of fluorescent proteins have been evolved, such as red fluorescent protein,[101,531−533] far-red fluorescent protein,[534] and yellow fluorescent protein.[535]

3.3. Metabolic Pathways

Directed evolution of entire metabolic pathways presents a greater challenge than directed evolution of single proteins or nucleic acids because the need to evolve multiple enzymes in a pathway, preferably simultaneously, increases the complexity of both library creation and library analysis. In addition, regulators and genetic elements associated with the pathway of interest may require optimization, which further expands the potential sequence space to be explored. Therefore, tailor-made methodologies are often employed for directed evolution of entire metabolic pathways. The optimum methods for genetic diversification and screening/selection of metabolic pathways are context-dependent and may vary depending on the target where it is utilized as an assimilation or biosynthetic pathway and available methods for screening/selection.

3.3.1. Assimilation Pathways. Assimilation pathways are particularly well-suited for directed evolution because cell growth and pathway efficiency are closely linked. Consequently, mutagenesis can be applied to the entire pathway followed by simple selection based on cell growth. Numerous attempts have been made to apply directed evolution to enhance the function of heterologous xylose assimilation pathways in *S. cerevisiae*. Alper and co-workers used the ICE method to engineer an isomerase-based xylose metabolic pathway consisting of xylose isomerase (XylA) and xylulose kinase (Xks1) (Figure 21A).[147] An isolate that emerged after 1 week of ICE contained a G164K substitution in Xks1, conferring a 21% increase in the exponential growth rate (Figure 21B). In a separate parallel experiment employing a variant xylose isomerase (XylA3*), two isolates were generated with mutations in XylA3*: the first with the mutation I433V and a silent nucleotide change (A1029G), and the second with three amino acid substitutions A48S, I433V and M435I, resulting in 14% and 16% improvements in the exponential growth rate, respectively (Figure 21C).

In contrast to the ICE method which targeted point mutations to a specific set of genes, an approach called "customized optimization of metabolic pathways by combinatorial transcriptional engineering (COMPACTER)" developed by Zhao and co-workers enables the generation of a massive combinatorial library of metabolic pathway variants consisting of multiple gene variants for each metabolic step expressed at varying levels.[536,537] This method was applied to a heterologous xylose utilization pathway in *S. cerevisiae* consisting of a xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulose kinase (XK), which led to a yeast strain with a xylose consumption rate of 0.92 g/L/h at a yield of 0.26 g ethanol per gram of xylose. A similar combinatorial method was used to create more than 8,000 xylose metabolic pathway variants, which led to a yeast strain that produced 0.31 g ethanol per gram of xylose and only 0.06 g per gram of xylose as a byproduct.[538]

Another combinatorial expression approach was used by Dueber and co-workers to enrich libraries of a three-gene xylose utilization pathway and an eight-gene xylose utilization pathway from *Scheffersomyces stipitis* under aerobic and anaerobic conditions.[539] A one-pot Golden Gate reaction was used to assemble backbones with plasmids containing expression cassettes driven by a set of five constitutive promoters of varying strength, resulting in a library containing all possible expression profiles of the three-gene and eight-gene pathways. The resulting library was then transformed into *S. cerevisiae* and enriched for the best performing xylose utilization pathways through serial cultures with xylose as a primary carbon source under both anaerobic and aerobic conditions. Unsurprisingly, the eight-gene xylose utilization pathway performed better than the minimally required three-gene xylose utilization pathway, and the best performing xylose utilization pathway under anaerobic conditions did not correspond to the best performing xylose utilization pathway under aerobic conditions.

In another example, Zhao and co-workers employed the COMPACTER method to develop a cellobiose-consuming *S. cerevisiae* strain with the highest efficiency reported at the time, yielding 0.44 g ethanol per gram of cellobiose.[536,537] Zhao and co-workers evolved the cellobiose utilization pathway by performing epPCR on the beta-glucosidase gene *gh1-1* and cellodextrin transporter gene *cdt-1* coupled with selection, which led to a 47% improvement in growth rate, 49% increase in cellobiose utilization rate, and 64% increase in ethanol productivity compared to the wild-type cellobiose utilization pathway.[540] The yield of ethanol from cellobiose was also slightly improved, from 0.41 ± 0.02 to 0.436 ± 0.004 g/g. By combining directed evolution with multigene pathway optimization, Zhao and co-workers developed a yeast strain capable of efficient cellobiose bioconversion, with a 6.4-fold increase in the cellobiose utilization rate and ethanol productivity.[541]

3.3.2. Biosynthetic Pathways. In contrast to directed evolution of assimilation pathways, where the function of the pathways can often be directly linked to the growth of the target organism or production of a growth-associated product, directed evolution of biosynthetic pathways can be substantially more difficult due to the required screening of the target product. Nonetheless, thanks to advances in tools for library creation and screening/selection, significant progress has been made over the past decades.

3.3.2.1. Organic Acids. Substantial efforts have been invested in microbial production of organic acids due to their high demand as building blocks for the synthesis of drugs, polymers, and other industrially important biochemicals. One of the target organic acids is 4-hydroxyphenylacetic acid, which has numerous uses in the synthesis of drugs and agrichemicals such as penicillin G and benzoprofen.[542] Using *E. coli* as a production host, Liu and co-workers identified the yeast Ehrlich pathway as an optimum biosynthetic route to 4-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

hydroxyphenylacetic acid. The component enzymes were evolved using epPCR, yielding two phenylpyruvate decarboxylase ARO10 and phenylacetaldehyde dehydrogenase FeaB variants that performed better than their wild-type counterparts. The expression of the evolved genes was then balanced using a tunable intergenic region sequence, improving production by 1.13-fold compared to the wild-type pathway. After implementing a quorum-sensing circuit to combat the toxicity of the heterologous pathway, the resulting strain achieved a maximum of 17.39 ± 0.26 g/L 4-hydroxyphenylacetic acid with a 23.2% molar yield, representing a further 46% improvement over the constitutively expressed evolved pathway.[542]

Glucaric acid is a key renewable chemical for planned replacement of petroleum-based polymers, such as polypropylene, polyethylene, nylon, and more.[543] Church and co-workers employed directed evolution to enhance biosynthesis of glucaric acid by coupling its production to individual cell fitness. This sensor-selector approach couples the sensory protein CdaR with the dual-selector TolC, which can be positively selected using sodium dodecyl-sulfate (SDS) and negatively selected with colicin E1. By combining this sensor-selector method with MAGE, Church and co-workers achieved a 22-fold increase in glucaric acid production using only a single round of diversification and selection. Although their maximum achieved titer of 1.2 mg/L glucaric acid is far lower than other reports, it was shown that the MAGE-competent *E. coli* K strain used for the directed evolution was not well-suited for glucaric acid production. A 300-fold increase in glucaric acid production was observed by expressing the same pathway in the *E. coli* BL21 strain.[543]

Muconic acid is another platform chemical important for production of renewable biopolymers.[544] Alper and co-workers engineered an *S. cerevisiae* strain with a biosensor responsive to aromatic amino acids as a surrogate for target pathway flux and then coupled this biosensor with an antimetabolite feeding strategy. Sequential rounds of directed evolution enriched strains with improved aromatic amino acid pathway flux, which was then redirected into muconic acid production by introduction of the target biosynthetic pathway, resulting in a maximum of 2.1 g/L muconic acid at a yield of 12.9 mg muconic acid/g glucose.[544]

3.3.2.2. Carotenoids. Carotenoids are red, orange, or yellow pigments which may possess vitamin A activity. Using a molecular breeding approach, Arnold and co-workers expressed shuffled phytoene desaturases and screened the resulting library for carotenoids, which led to a novel pathway for production of 3,4,3′,4′-tetradehydrolycopene in addition to other related carotenoids.[545] Lycopene, a bright red carotenoid molecule, has been a popular target for metabolic engineering and directed evolution. Church and co-workers employed MAGE for accelerated evolution of a lycopene pathway and isolated a strain with more than 5-fold improvement in lycopene production.[104] Through triclosan-induced chromosomal evolution, Liu and co-workers developed a lycopene-producing *E. coli* strain free of plasmids or antibiotic markers capable of accumulating 33.43 mg lycopene per gram of dry cell weight.[546]

3.3.2.3. Other Examples. In pursuit of an optimized pathway for the biosynthesis of amorpha-4,11-diene, a precursor to the antimalarial drug artemisinin, Keasling and co-workers subjected a heterologous *S. cerevisiae* mevalonate pathway expressed in *E. coli* to directed evolution by generating libraries of tunable intergenic regions (TIGRs).[547] TIGRs can be used to vary the relative expression of two reporter genes over a 100-fold range and simultaneously tune the expression of several genes within an operon. A 7-fold increase in mevalonate production was achieved by employing a library of TIGRs to balance the expression of the mevalonate biosynthesis pathway.[548]

Salis and co-workers aimed to enhance the cellular supply of NADPH by transferring the five-enzyme Entner−Doudoroff pathway from *Zymomonas mobilis* into *E. coli* and then enhancing its activity by accelerated evolution via MAGE to enable rapid regeneration of NADPH. After creating a library of 10^6 pathway variants via MAGE, an NADPH-dependent blue fluorescent protein (mBFP) was employed to quickly screen for 624 variants with high NADPH regeneration rates. Among these, 22 strains were subjected to in-depth characterization of *in vivo* NADPH regeneration rates and NADPH-dependent biosynthesis rates to identify potential relationships between enzyme expression levels and NADPH regeneration. However, higher expression levels were not consistently linked with the highest NADPH regeneration rates. The best variant identified exhibited a 25-fold increased NADPH regeneration rate as measured by the mBFP fluorescent reporter and when combined with an optimized terpenoid pathway, the evolved Entner−Doudoroff pathway increased terpenoid titer by 97%.[549]

Using the same sensor-selector strategy described above, Church and co-workers enhanced production of the flavonoid naringenin in *E. coli*. Instead of focusing directed evolution on the naringenin biosynthetic pathway, efforts were made toward genomic targets that can enhance the supply of the naringenin precursors tyrosine and malonyl-CoA. A toggled selection approach was employed, wherein those cells that gain mutations enabling survival despite not producing the target chemical are selectively killed in each round of evolution. Based on the naringenin sensor TtgR, To1C dual selection was used to couple toggled selection with MAGE to generate diversity followed by selection of around 20 colonies from each round of directed evolution and enrichment. Four rounds of directed evolution led to isolation of an evolved strain with 36-fold higher naringenin production compared to the parental strain.[543]

In another example, Church and co-workers used CoS-MAGE to improve the production of aromatic amino acid derivatives in *E. coli*. Specifically, CoS-MAGE was used for simultaneous and combinatorial insertion of multiple T7 promoters into 12 genomic operons, generating 80 unique variants from the fifth iteration of CoS-MAGE for further characterization. Tryptophan was indirectly measured by using plasmid-based expression of a *Methylophaga sp.* flavin-containing monooxygenase followed by extraction and quantification of indigo and indirubin pigments. The best performing variant produced over 8.6 mg indigo per gram of dry cell weight, a 4-fold improvement over the parental strain.[550]

Although there are numerous potential methods to introduce genetic diversity into a metabolic pathway, the large sequence space generated necessitates effective screening methods for successful application of directed evolution toward optimization of entire metabolic pathways. The highly successful efforts highlighted above directly linked pathway efficiency to cellular fitness, either by evolving a pathway essential for the primary available carbon source[147,536−538] or

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

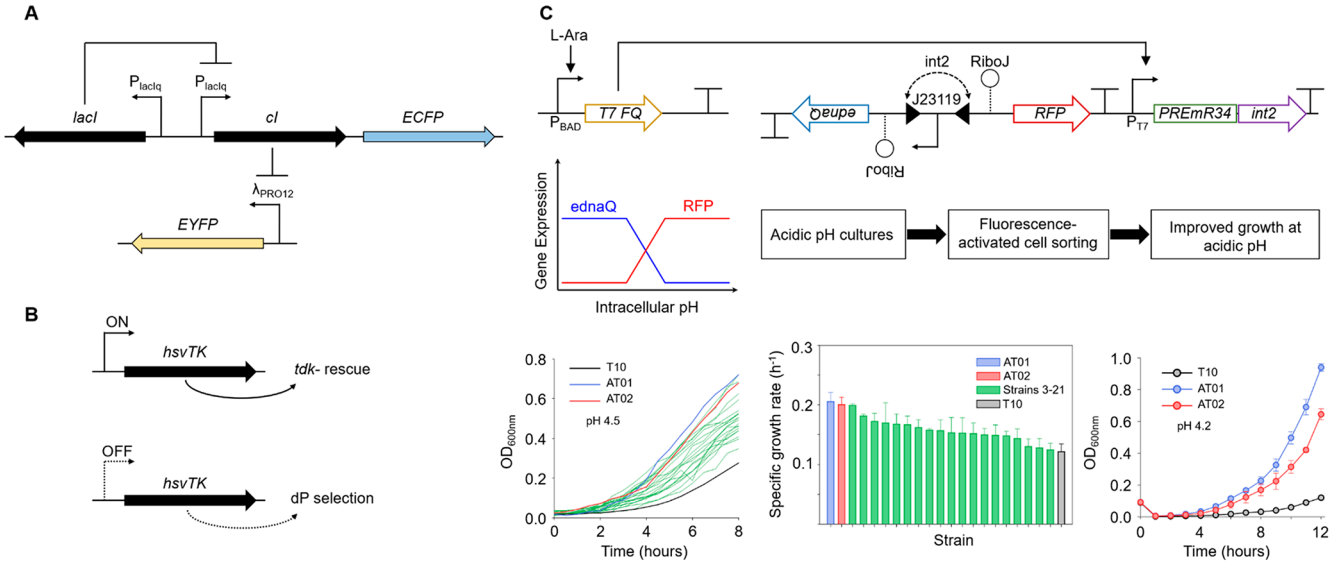


Figure 20. Selected examples of directed evolution of genetic circuits. (A) Circuit designed by Yokobayashi and co-workers encoding an IMPLIES logic gate.[554] EYFP is inversely linked to the concentration of IPTG via cI-based repression of the EYFP-driving λ_{PRO12} promoter, which is in turn repressed by lacI in the absence of IPTG. (B) Optimized dual selection strategy developed by Umeno and co-workers employ the *hsvTK* gene for positive selection of the *tdk*- phenotype and negative selection using dP.[555] (C) Genetic circuit harnesses arabinose (L-Ara) induction of riboswitch sensor (*PREmR34*) to selectively drive expression of an integrase (*int2*). At low intracellular pH, int2 flips the orientation of the J23119 promoter, turning off expression of a mutagenic dnaQ variant (*ednaQ*) and switching on expression of a fluorescent reporter (*RFP*). This allows selection of desirable phenotypes using FACS and enrichment of strains capable of growth at low pH. From the parental T10 strain (black), a variety of improved strains (green) were screened at pH 4.2. Among these, the evolved strains AT01 (blue) and AT02 (red) exhibited substantially improved growth down to pH 4.2.[365] (Adapted with permission from ref 365. Copyright 2017 Springer Nature. Creative Commons Attribution 4.0 International License. To view a copy of license, visit https://creativecommons.org/licenses/by/4.0/. Original Figures 4a,b,d,e and 5b were adapted in part C.)

by linking a biosensor for the target product to survival of the cell.[543] Producing a pigment, such as carotenoids, simplifies screening as colonies of successfully evolved variants can be visually identified on a plate among thousands of nonfunctional background isolates.[545,546] Efforts to apply directed evolution to entire metabolic pathways would be greatly aided by the development of innovative new screening strategies or generalized methods to select for producers of specific target molecules.

3.4. Genetic Circuits

Preprogrammed responses and complex behaviors can be implemented in engineered organisms using genetic circuits, but fine-tuning circuit function can be an immense task due to the wide range of factors linking the input signal to the desired circuit response. While design and implementation of a novel genetic circuit is an inherently rational process, directed evolution is an invaluable tool for optimizing circuit function. Arnold and co-workers rationally designed a nonfunctioning genetic circuit encoding the IMPLIES logic function, which was comprised of separate NOT and OR logic functions. Ultimately, the circuit shown in Figure 20A inversely links the concentration of Enhanced yellow fluorescent protein (EYFP) to the concentration of the external inducer IPTG, by repressing EYFP expression with the cI protein, which is in turn repressed by LacI in the absence of IPTG. For this IMPLIES logic statement to function properly, the distinct NOT and OR statements must produce biological outputs, such as protein concentrations, which are in ranges compatible with each other and can be interpreted as the same logic level. While this matching is not easy to achieve via rational design, directed evolution can be used to generate a functional circuit from a nonfunctioning initial rational design. Variant libraries of the cI protein and its RBS were generated followed by evolution for optimal circuit function by first screening for high amounts of the fluorescent protein in the absence of IPTG, then for low amounts of the fluorescent reporter in the presence of IPTG. A wide range of cI variants were identified which enabled a functional circuit: truncations of the C-terminal domain, multiple mutations in the coding sequence, and modifications to the RBS.[551] Similarly, Voigt and co-workers constructed a circuit which required tuning the linkage between a sensor output gene.[552] A library of random ribosome binding sites was constructed and screened for selective induction under anaerobic growth or arabinose induction.

While the first report of directed evolution of a genetic circuit was fairly specific to the IMPLIES logic statement, Arnold and co-workers later expanded upon this technique with the development of a dual selection strategy that allows selection of ON and OFF states by coupling the circuit output with the survival or death of the host strain.[553] Using this method, a generic selection method can be connected to any genetic circuit that results in ON/OFF gene expression as an output. The dual selection strategy was implemented using the genetic circuit shown in Figure 20: the ON state drives polycistronic expression of *tetA*, conferring tetracycline resistance, and *bli*, encoding a protein inhibitor of β-lactamase, while a constitutive promoter expresses *bla* encoding β-lactamase. When this circuit is in the ON state, the cell population is resistant to tetracycline and susceptible to carbenicillin. When in the OFF state, *tetA* and *bli* are no longer expressed, causing susceptibility to tetracycline and resistance to carbenicillin. This circuit was then placed under the control

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

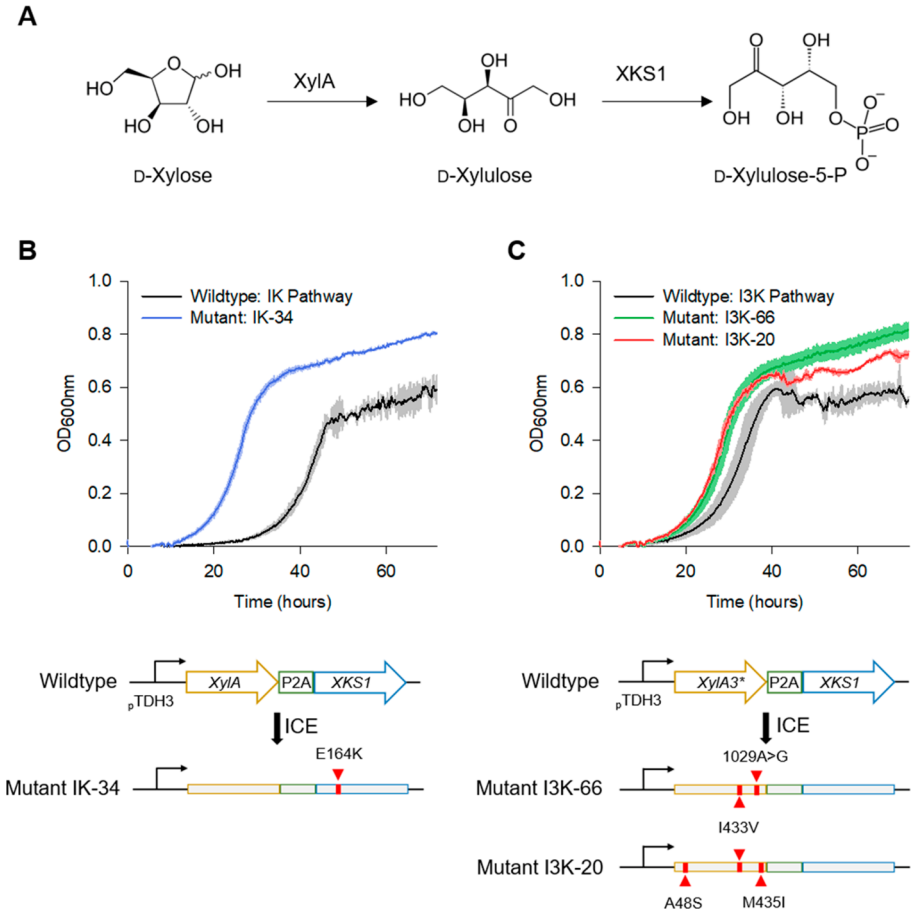


Figure 21. *In vivo* continuous evolution (ICE) of the xylose metabolic pathway. (A) The isomerase-based xylose metabolic pathway employs xylose isomerase XylA to convert D-xylose to D-xylulose followed by phosphorylation via XKS1 into D-xylulose-5-phosphate. (B) A single expression cassette consisting of the TDH3 promoter driving expression of XylA and XKS1 linked by a P2A subjected to rounds of ICE generated a single E164 K mutation in the XKS1 gene, improving the growth rate by 21% and reducing the lag phase by 18 h. (C) Employing ICE on an expression cassette employing the variant xylose isomerase XylA3* generated two improved variants with 6 h shorter lag phases: I3K-66 with an I433 V and silent 1029A > G mutation improving growth rate by 14% and I3K-20 with A48S, I433 V and M435I mutations exhibiting a 16% improved growth rate.[147] (Adapted with permission from ref 147. Copyright 2016, Springer Nature. Creative Commons Attribution 4.0 International License. To view a copy of license, visit https://creativecommons.org/licenses/by/4.0/. Original Figure 5c,d was adapted in parts B and C.)

of the λPRO12 promoter and expressed alongside LacI so that the presence of IPTG resulted in an OFF circuit (Figure 20). This circuit was mixed with cells containing nonfunctional background circuits at a molar ratio of 200:1 and enriched via selecting for tetracycline resistance without IPTG, then carbenicillin with IPTG. A single round of selection resulted in 155-fold enrichment of the circuit plasmid. Yokobayashi and co-workers later simplified this approach by allowing selection of ON/OFF circuits using the single selection marker *tetA*.[554] By taking advantage of the $NiCl_2$ sensitivity of *tetA*-expressing cells, both positive and negative selection can be accomplished using this single marker. Three rounds of directed evolution using this simplified circuit enabled selection of the desired switch from a 2,000-fold excess background of nonfunctional switches.

Umeno and co-workers further improved the dual selection methodology by employing the Herpes simplex virus thymidine kinase (hsvTK), which induces cell death in the presence of the mutagenic nucleoside dP and acts as an excellent OFF-selector (Figure 20B).[555] Thymidine kinases can also rescue *tdk*- strains from thymidine deficiency, enabling ON selection as well. Due to the impressive selection strength of the hsvTK system, successive ON/OFF selection rounds can be accomplished at a selection efficiency of over 10^4 per round. Recently, Umeno and co-workers employed hsvTK dual selection to evolve a highly stringent LuxR *Vibrio fischeri* quorum sensor.[556] Stringency is often generated by successive ON/OFF selection rounds, aiming to generate high induced expression and low leakiness, respectively. Interestingly, four successive rounds of ON-selection alone generated a population of variants which required ligand binding for proper protein folding, nearly eliminating leakiness in the absence of the induction ligand.

Throughout these efforts, cellular fitness was an essential prerequisite for optimization of the genetic circuit by directed evolution. To decouple cellular fitness from optimization of circuit function, Ellington and co-workers developed compartmentalized partnered replication (CPR), which combines *in vivo* protein expression and *in vitro* PCR amplification steps.[557] The six amino acid residues of the T7 RNA polymerase specificity loop were fully randomized and used to drive the modified T7 promoter P_{CGG}. Following 16 rounds of directed evolution and amplification, a variant RNA polymerase was identified with activity toward P_{CGG} at a similar strength as the wild-type polymerase toward the wild-type promoter, but with less than 0.1% cross-reactivity with the wild-type T7 promoter.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

CPR was also shown to be applicable to coevolution of entire genetic circuits. An aminoacyl-tRNA synthetase:tRNA pair from *S. cerevisiae* was coevolved to enable incorporation of the unnatural amino acid 5-hydroxy-L-tryptophan (5HTP) by suppressing amber codons. After ten rounds of CPR, variants were transformed with GFP containing three amber codons and assayed for their ability to produce a fluorescent output. The CPR-evolved 40A tRNA variant together with the CPR-evolved 5HTP incorporating 5OH-R3−13 tRNA synthetase gave a robust fluorescent output by efficient incorporation of 5HTP in response to the GFP amber codons.[557,558]

In a separate approach employing riboswitch-based sensors for pH-dependent control of gene expression, Chang and co-workers used a genetic circuit to evolve acid-tolerant phenotypes by programming cells with an inverse relationship between the *in vivo* mutagenesis rate and the intracellular pH.[365] This was accomplished by constitutive expression of *ednaQ*, an error-prone variant of *dnaQ*, the ε subunit of *E. coli* DNA polymerase III which contributes to the proofreading capacity of the replication machinery, which on its own led to a 100-fold increase in genome replication error rate. Using a pH-sensitive riboswitch, cells capable of buffering against acidic extracellular pH conditions trigger expression of an integrase, which inverts the *ednaQ* promoter, switching off its expression while activating expression of a fluorescent reporter gene, permanently marking the cell as having evolved a desirable phenotype (Figure 20C). Three rounds of directed evolution led to a fluorescent population from which two strains were isolated with substantially improved acid tolerance as compared to the parental *E. coli* strain. The evolved strains AT01 and AT02 exhibited increased growth in low pH conditions and a substantial buffering of their internal pH when in the presence of several industrially relevant organic acids.

Design and evolution of a genetic circuit can be limited by the variety, quality, and orthogonality of component parts. In pursuit of a large set of programmable and orthogonal circuitry, Ellington and co-workers employed CPR to generate a synthetic phylogeny of Trp repressors.[518] TrpR variants were evolved with activities toward novel operator sequences in addition to altered ligand binding sites, enabling regulation by L-tryptophan or several analogues onto a variety of sequence options. Moreover, tethering these synthetic repressors enables generation of intramolecular protein logic by linking different functions into a single protein.

Directed evolution is an essential tool for generating component parts and optimizing the function of genetic circuits. However, current implementations are severely limited by selection/screening capabilities and thus have largely focused on relatively simple circuits.[559] Because of the vast complexity of selection, new tools and techniques will need to be developed for directed evolution of circuits with outputs having spatial[560] and temporal[561,562] variations. Implementing these circuits in cell consortia adds even further complexity as evolving cell−cell communication networks and population-level behaviors[563] will require the generation of vast libraries and precision selection techniques. Most likely, the most fruitful efforts will combine rational design with directed evolution: rational design of complex circuit networks based upon logic gates and component proteins optimized through directed evolution.

3.5. Viruses

Viruses have short generation time and their capacity to adapt to new hosts and environments is highly dependent on the ability to generate diversity in a short of time.[564] Mutation rates vary among viruses: RNA viruses usually mutate faster than DNA viruses, single-stranded viruses mutate faster than double-strand viruses, and genome size appears to correlate negatively with mutation rate.[564,565] Viral mutation rates are modulated by different factors such as polymerase fidelity, replication mode, genomic architecture, and access to postreplicative repair.[564,565] Viruses evolve through changes in their genomic RNA (or DNA) under selective pressure, and the best adapted variants will be quickly enriched, spread, and continue their contagion. Because many viral infections have a profound impact on host fitness and survival and thus host evolution, understanding viral genome evolution can provide insights about potential drug and vaccine targets.[566,567] Directed evolution has proved to be a powerful strategy for viral genome evolution. Since the mechanisms and processes have been comprehensively reviewed elsewhere,[564,565,568−571] we will mainly focus on two main applications: driving nonviral protein expression and optimizing gene therapy.

3.5.1. Driving Nonviral Protein Expression. The directed virus evolution strategy can be used to improve gene expression systems by making them an integral and essential part of the virus,[572] thereby optimizing nonviral protein expression. Bamford and co-workers developed an *in vitro* evolution system based on a recombinant double-stranded RNA bacteriophage, φ6 to monitor the trajectory of molecular changes in RNA genome.[573] The recombinant φ6, containing an integrated β-lactamase gene (*bla*) marker was propagated in carrier-state bacterial cells in the presence of β-lactam antibiotic (cefotaxime). After several passages, the carrier-state bacterial cells showed resistance to high concentrations of cefotaxime, and three nonsynonymous mutations were found in β-lactamase, two of which (E104 K and G238S) have been previously reported for β-lactamases from cefotaxime-resistant bacterial isolates. In a similar study, Das and co-workers optimized the tetracycline-regulated gene expression system (Tet system) in mammalian cells by incorporating the components of the Tet system in the human immunodeficiency virus (HIV)-1 virus.[572] Upon long-term replication of HIV-reverse transcriptional activator (HIV-rtTA) virus in human T cells, a virus variant with single amino acid substitution (F86Y) was obtained that enhanced virus replication potential, and thus transcriptional activity and doxycycline (dox) sensitivity of rtTA. An rtTA variant was generated with 5-fold higher activity at high dox levels and 25-fold higher sensitivity to dox compared to the wild-type rtTA. Such a variant would be of particular advantage when the HIV-rtTAvirus is used as an anti-HIV-1 vaccine because of the reduced risk of unintended reactivation of the virus after vaccination. In another study, a synthetic system was designed with a genetically engineered *E. coli* and phage M13 virus.[574] *E. coli* provides nutrients and energy necessary for virus propagation, while phage M13 enhances infected hosts to survive under fatal concentration of antibiotics. The system was then applied to evolve a virus-carried heterologous gene (*luxR*) under increasing selective pressure, which consequently allowed the infected hosts to survive against antibiotics. As a result, several evolved LuxR variants with improved activity (up to 17-fold) were obtained from a limited number of viruses randomly isolated after a small number of generations. This

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

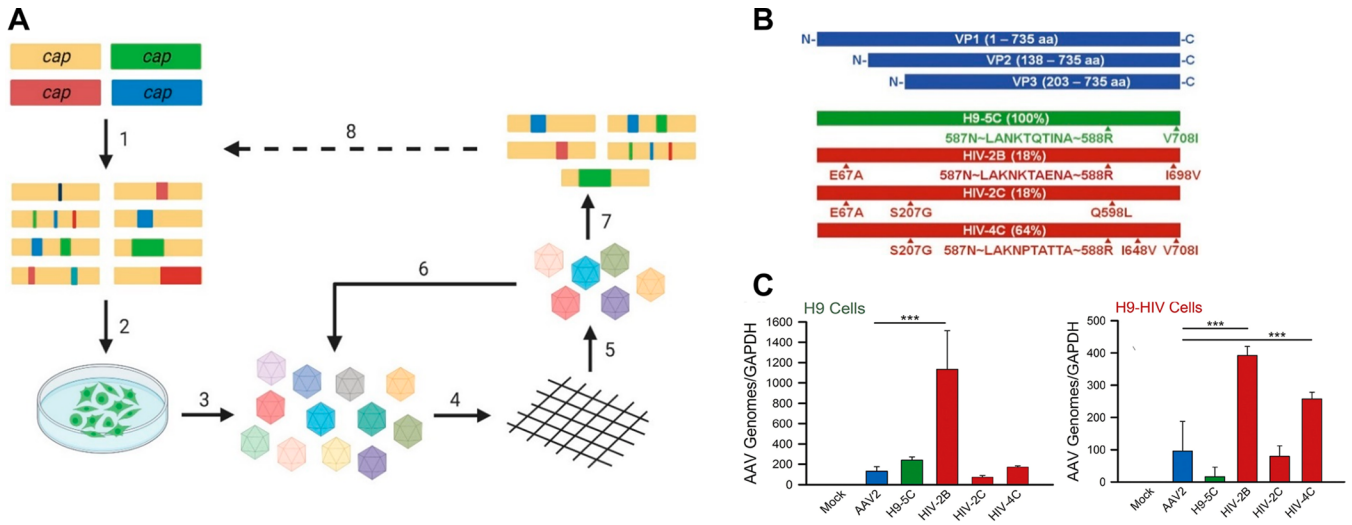


Figure 22. Directed evolution of AAV and its application in selecting for novel AAV capsids on an HIV-1 producer T cell line. (A) Workflow for direct evolution of AAV. Following the arrows: (1) a viral library is created by mutagenesis and recombination of the *cap* gene; (2) viruses are packaged, typically in HEK293 cells using plasmid transfection, resulting each virion contains the *cap* gene encoding that virion's capsid protein; (3) viruses are harvested and purified; (4) the viral library is introduced to selective pressure; (5) successful viruses are amplified and recovered using adenovirus rescue; (6) successful clones are enriched through repeated selection steps; (7) sequencing to confirm the isolated viral DNA with selected *cap* genes; (8) selected *cap* genes are mutated again to serve as a new starting point for further selection steps to iteratively increase viral fitness.[576] (Adapted with permission from ref 576. Copyright 2012 Springer Nature. Original Figure 1a was adapted in this Figure 22A.) (B) Novel AAV variants after six rounds of selection. Wild-type AAV encodes three viral capsid proteins (VP1, VP2, and VP3; blue). After six rounds of selection, the AAV library converged to a small number of sequences. One predominant sequence for H9 cells (green) and three predominant sequences for H9−HIV cells (red) are shown with percentages indicating prevalence within the population. (C) Number of cell-associated genomes for selected AAV variants. Specific AAV variants were packaged and used to infect either H9 or H9−HIV cells. qPCR was performed on total cell DNA isolated from the infected cells or control cells with normalization against the GAPDH gene.[581] (Reproduced with permission from ref 581. Copyright 2017, Elsevier. Original Figures 3 and 4 were reproduced in parts B and C.)

synthetic system offers an *in vivo* targeted evolution strategy capable of facilitating evolution through artificially created interspecies cooperation.

3.5.2. Optimizing Gene Therapy. Gene therapy vectors based on adeno-associated virus (AAV) are the leading platforms for the treatment of a variety of human diseases. Efforts have been made to increase preclinical and clinical successes in AAV-mediated gene therapy by developing desirable AAV capsids, optimizing genome designs, and harnessing novel revolutionary strategies.[575] Since the virology, fundamentals, mechanisms, vector designs, and biosafety of AAV have been comprehensively reviewed,[575−579] here we outline some processes in creating novel classes of AAV variant libraries for diverse applications using directed evolution strategies.

Directed evolution enables relatively rapid selection and isolation of AAV variants with novel and therapeutically valuable properties (Figure 22A).[576,580] It was applied to isolate AAV capsids with increased tropism toward an HIV-1 producer T cell line.[581] A library of AAV variants was used to infect HIV-1-infected and uninfected H9 T cells followed by AAV amplification with wild-type. Six rounds of selection were performed, and the H9 T cells were successfully infected with all three wild-type viruses (AAV, adenovirus, and HIV-1). Four resulting variant capsids (H9-5C, HIV-2B, HIV-2C, and HIV-4C) (Figure 22B) were used to package an AAV vector and the corresponding AAV variants were subsequently used to target parental and HIV-1 infected H9 T cells. A quantitative polymerase chain reaction (qPCR) assay revealed that two (HIV-2B and HIV-4C) of the four capsid variants showed a significant increase in the amount of cell-associated viral DNA compared to wild-type viruses (Figure 22C). This study offers a potential gene therapy approach with the goal of delivering an anti-HIV-1 therapeutic molecule such as a small interfering RNA molecule or a genome editing system.

Inherited retinal degenerative diseases are a clinically promising target of AAV-mediated gene therapy. Flannery, Schaffer, and co-workers implemented *in vivo* directed evolution to engineer AAV variants capable of delivering the gene cargo to the outer retina from the vitreous.[582] An evolved AAV variant (7m8) mediated highly efficient delivery to all retinal layers in mice and nonhuman primates. It also mediated therapeutic gene delivery to photoreceptors in two mouse models of retinal disease, enabling noninvasive, long-term histological and functional rescue of disease phenotypes across the entire retina. These findings have important clinical implications for the development of gene therapies for X-linked retinoschisis, Leber's congenital amaurosis, and other retinal diseases requiring robust, panretinal gene expression without retinal detachment. Their recent study introduced deep sequencing to guide *in vivo* directed evolution of AAV and used a GFP-barcoded library to select high-performing AAV vectors.[583] Six rounds of *in vivo* selection in primates resulted in vectors with redirected tropism to the outer retina and increased gene delivery efficiency to retinal cells.

In addition to AAV vectors, recently developed synthetic virus-like nucleocapsids from nonviral protein cages have great potential for the delivery of nucleic acids, which may avoid the safety risks and engineering challenges associated with viruses.[584,585] The ability of synthetic nucleocapsids, previously designed icosahedral protein assemblies with positively charged inner surfaces, was explored to evolve virus-like properties, for example packaging their own full-length mRNA genomes.[584] The diversified populations or variants were

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Table 6. Selected Examples of Whole Cell Engineering by Directed Evolution Strategies

applications	compounds	methods	genetic modifications	improvements	ref
Substrate Utilization	Xylose	Three rounds of RAGE in *S. cerevisiae*	Overexpression of *VPS13*, *MDH1*, and *COX5A*; down-regulation of *CDC11*	29% faster xylose utilization, and 45% higher ethanol productivity	588
	Galactose	One round of genome-wide perturbation in *S. cerevisiae* and 10 times serial subcultures	Overexpression of previously confirmed *SEC53* and novel targets *SNR84* and *TUP1*	250% higher in both galactose consumption rate and ethanol productivity for the truncated TUP1 overexpression	592
	Glycerol	Four rounds of genome shuffling in *Clostridium diolis*	N/A	80% higher starting glycerol concentration for variant 98; and 80% higher 1,3-propanediol yield in variants GSHM2, 4	590
Chemical Production	Tylosin	Two rounds of genome shuffling in *Streptomyces fradiae*	A single-nucleotide mutation in a regulatory gene within the tylosin biosynthetic pathway	9-fold more tylosin than the starting strain	88
	Lycopene	5−35 cycles of MAGE in *E. coli*	24 genes in the DXP pathway were modified, and their RBS sequences were optimized	5-fold increase in lycopene production	104
	Carotenoids	5 iterative cycles of SCRaMbLE (MuSIC)	yJBD069 strain was generated from yJBD057 strain, while yJBD057 had a much larger duplication (130 kb) from *YEL072W* to *YEL012W* and a larger deletion in synIII (*YCL073C-YCL009C*) than those in yJBD048	38.8-fold increase of carotenoids production	119
Tolerance/ Resistance	Furfural	Two rounds of RAGE in *S. cerevisiae*	*SIZ1* deletion or knockdown	73% higher maximum growth rate of *siz1*Δ strain than that of wild-type in the presence of 0.8 g/L furfural	595
	Lactic acid	Five rounds of genome shuffling in *Lactobacillus*	N/A	The shuffled strain F5 could grow at pH 3.8, and produced >5 g/L lactic acid, 3-fold higher than the wild-type strain at pH 4.0	89
	Isobutanol/1-butanol	One/two rounds of tunable RNAi in *S. cerevisiae*	Hsp70 family as a key regulator of isobutanol tolerance in a single round of screening; the alcohol dehydrogenase and enolase were identified through two rounds of screening	Downregulation of Hsp70 family genes increased cell growth rate by 64% in 12 g/L isobutanol; the combined downregulation of *ADH* and *ENO* improved growth up to 3,100% in 10 g/L 1-butanol	598

generated using *E. coli* as an expression host. After several generations of directed evolution, the markedly improved genome packaging efficiency (more than 133-fold), stability in blood (from less than 3.7% to 71% of packaged RNA protected after 6 h treatment), and *in vivo* circulation time (from less than 5 min to ∼4.5 h) were achieved. The resulting synthetic nucleocapsids were able to package one full-length RNA genome per 11 icosahedral assemblies, comparable to the best recombinant rAAVs. A nonviral protein cage formed by *Aquifex aeolicus* lumazine synthase and its encoding mRNA was converted to virus-like nucleocapsids by directed evolution.[585] The evolved capsid proteins specifically recognize packing signals on cognate mRNAs, and these RNA tags can also be used to direct preferential packaging of other genes. These studies demonstrate that there are simple evolutionary paths through which protein assemblies can acquire virus-like genome packaging and protection capabilities. As such, these systems have great potential for therapeutic nucleic acid delivery and vaccine development.

3.6. Whole Cells

Directed evolution has emerged as a powerful tool for engineering various organisms ranging from bacteria to higher eukaryotes on the genome level. It can not only facilitate selection of desired phenotypes but also help researchers explore the genetic targets that are responsible for these phenotypes and unlock nature's biosynthetic potentials.[586] In this section, we highlight several directed evolution strategies on a genomic scale and their applications in improving substrate utilization, chemical production, cell tolerance/resistance, and genotype-phenotype mapping.

3.6.1. Improving Substrate Utilization. A direct link exists between cell fitness and substrate assimilation capacity. Efficient utilization of substrates, especially the undesirable carbon sources, has become one of the most essential objectives for constructing microbial cell factories. Directed genome evolution can uncover genomic variants conducive to robust growth on substrates of interest.

Xylose is a major component of lignocellulosic biomass (one of the most abundant feedstocks), and its efficient utilization is crucial for lignocellulosic biofuel production. To improve xylose fermentation in *S. cerevisiae*, DNA shuffling was used to recombine the whole genome of *Pichia stipitis* that has high xylose-fermenting capability with that of *S. cerevisiae*.[587] After two rounds of genome shuffling and screening, one evolved *S. cerevisiae* strain ScF2 was able to utilize high concentrations of xylose (100 g/L to 250 g/L) for ethanol production, and the ethanol producing rate is even higher than that of *P. stipitis*. In another study, RNAi-assisted genome evolution (RAGE) was performed in *S. cerevisiae* to improve xylose utilization.[588] In particular, a full-length cDNA library was constructed in the *S. cerevisiae* SR8 strain and processed into genome-wide modulation cassettes encoding both overexpression and down-regulation mutations. In each round of RAGE screening, modulation cassettes conferring the largest phenotypic improvement were integrated into the genome to create new parent strains for the next round to identify potential synergy between beneficial mutations. After three rounds of RAGE, overexpression of *VPS13*, *MDH1* and *COX5A* as well as down-regulation of *CDC11* were found to improve the xylose utilization rate and ethanol productivity by 29% and 45%, respectively (Table 6). When applied to yeast harboring synthetic chromosome V, SCRaMbLE could generate strains with dramatically enhanced cell growth rates in xylose medium.[118] The average growth rate of resultant strain XD4-pJCH006 was 0.2 h^{-1} in baffled flask containing xylose, which compares favorably to the 0.18 h^{-1} maximum growth rate

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

reported for similar conditions after extensive optimization of the xylose utilization pathway.[589]

In addition to xylose, utilization of other substrates was also enhanced by different directed genome evolution methods. Examples include glycerol utilization by genome shuffling in *Clostridium diolis*,[590] CRISPR-assisted RAGE in *S. cerevisae*,[591] galactose fermentation by genome-wide perturbation in *S. cerevisiae*,[592] and fermentation of sugars in hardwood spent sulfite liquor (HWSSL) by genome shuffling in *S. cerevisiae* (Table 6).[593]

3.6.2. Enhancing Chemical Production. Maximizing the production of chemicals is a challenging metabolic engineering task thanks to the underlying conflict between fast, steady cell growth, and repurposing of cellular resources to produce a high titer of desirable compounds.[594] Directed genome evolution can be used to overcome this conflict and optimize the production of chemicals.

Tylosin, a complex polyketide antibiotic is commercially biosynthesized by *Streptomyces fradiae*.[88] Through two rounds of genome shuffling, seven strains were identified that produce higher tylosin titers than the industrial SF21 strain (derived from natural isolate SF1 via 20 cycles of classical strain improvement) under high-throughput conditions.[88] The best two shuffled strains, GS1 and GS2, produced up to 9-fold more tylosin than SF1 and were statistically the same as SF21 (Table 6). Genome sequencing showed that SF21, GS1 and GS2 all had a single-nucleotide change in a regulatory gene, *tylQ* compared to SF1. However, the mutation in GS1 and GS2 was at a different position from that in SF21.

1-Deoxy-D-xylulose-5-phosphate (DXP) is an important intermediate in the nonmevalonate (or mevalonate independent) pathway, and the optimized DXP biosynthesis pathway can be used to produce isoprenoid compounds of industrial and pharmaceutical relevance, such as lycopene.[104] MAGE was applied to optimize the DXP biosynthesis pathway for overproduction of lycopene in *E. coli*.[104] Twenty-four gene targets in the DXP pathway were modified simultaneously using a complex pool of synthetic DNA, generating 4.3 billion combinatorial genomic variants per day. Several isolated variants exhibited 5-fold increase of lycopene production within 3 days; the highest lycopene yield of ∼9000 ppm (μg per g dry cell weight) is better than documented yields (Table 6).

Besides tylosin and DXP, the production/biosynthesis of many other valuable compounds have been greatly improved using various directed genome evolutions strategies. Examples include the enhanced biosynthesis of violacein and penicillin G by SCRaMbLE in *S. cerevisiae*[118] and the optimized carotenoid pathway by MuSIC (Multiplex SCRaMbLE Iterative Cycling) (Table 6).[119]

3.6.3. Increasing Tolerance/Resistance. Under harsh growth conditions or in the presence of stressors, cells can activate stress responses by reallocating cellular resources and thus inhibit their growth potential. Directed genome evolution has been established as an efficient tool to overcome the growth inhibition induced by various stressors such as low pH/acid, growth inhibitors, and toxic products.

Furfural is one of the major inhibitors present in lignocellulosic hydrolysates, which is formed from the dehydration of pentoses during dilute acid pretreatment of lignocelluloses.[595] The toxicity of hydrolysates correlates with furfural concentration, with less than 5 g/L of furfural completely inhibiting cell growth and significantly reducing the yield and productivity of desired products.[595] Therefore, increasing furfural tolerance in microorganisms should provide a cost-effective way for lignocellulose fermentation. Using genome-wide RNA-interference (RNAi, in a design similar to RAGE described in section 3.6.1), Zhao and co-workers identified the E3 SUMO-protein ligase gene *SIZ1* as a novel determinant of furfural tolerance in *S. cerevisiae*.[595] The *siz1*Δ cells exhibited improved furfural tolerance, which was accompanied by rapid furfural reduction to furfuryl alcohol and resulted in >2.5-fold higher ethanol productivity in the presence of 0.8 g/L furfural (Table 6). In addition, genome shuffling was applied to *Z. mobiliz* for enhanced furfural tolerance.[596] After two rounds of directed evolution, the evolved strains exhibited improvements of 76.8% in cell growth rate and 88.5% in ethanol yield compared to the parental strains in the presence of 5 g/L acetic acid (another major inhibitor in lignocellulosic hydrolysates) and 3 g/L furfural.

In addition to furfural, other stressors can also be relieved through directed genome evolution. Examples include improving lactic acid tolerance in *Lactobacillus* by genome shuffling (Table 6),[89] improving acetic acid tolerance in *S. cerevisiae* by RAGE,[597] and increasing isobutanol and 1-butanol tolerance in *S. cerevisiae* using a tunable RNAi screening approach (Table 6).[598]

3.6.4. Genotype−Phenotype Mapping. Adaptive evolution is a useful strategy to engineer a microorganism for a desired phenotype, but it is challenging to efficiently link genotype to phenotype. Predicting the functional consequences of genetic variation is one of the fundamental challenges in understanding phenotypic diversity and engineering desirable traits for biotechnological applications.[132] Facile CRISPR mediated HDR-assisted genome-scale perturbations mentioned in section 2.1.2.3 provide ways to obtain variants with the desired phenotype and identify corresponding genes contributing to those changes, and thus facilitate the subsequent engineering process. For example, Zhao and co-workers developed trackable genome wide knockout libraries of *S. cerevisiae* using CHAnGE and identified double variants SIZ1ΔLCB3Δ and BUL1ΔSUR1Δ with 40- and 20-fold improvements in furfural and acetate tolerance, respectively. Those gene targets had not been reported before.[130] More recently, the same group reported another whole genome-wide engineering tool, MAGIC, that provides more comprehensive genome engineering through a triple-mode of genome-scale activation, interference, and deletion.[668] In total, 37,817, 37,870, and 24,806 unique guide sequences, covering more than 99% of ORFs and RNA genes, were synthesized for the CRISPRa, CRISPRi, and CRISPRd libraries, respectively. MAGIC can identify previously uncharacterized genetic determinants of complex phenotypes, especially those having synergistic interactions when perturbed to different expression levels; for instance, the SIZ1i-NAT1a-PDR1i combination exhibited cell growth under YPD medium with 30 mM furfural, and HOC 1d-NUP157i improved yeast surface display capacity of recombinant proteins by 3-fold.

In *E. coli*, a tractable, double-stranded DNA cassette mediated homologous recombination strategy (TRMR) was developed to modify the expression of >95% of *E. coli* genes in a single day, and map thousands of genes affecting cell growth in various media and with the growth inhibitors within 1 week.[599] For example, through TRMR, the *xylA* up allele, which causes overexpression of *xylA* and *xylB*, was shown to confer growth in the presence of D-fucose, a nonmetabolizable

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

analog of arabinose that inhibits the ability of *E. coli* to use arabinose as a carbon source. This also suggests that *E. coli* xylose isomerase (encoded by *xylA*) may have *in vivo* L-arabinose isomerase activity. In a related study, TRACE (tracing combinatorial engineered libraries) enables the simultaneous mapping of millions of combinatorially engineered genomes at single-cell resolution.[600] It was used to identify genotype-to-phenotype correlations and track the evolutionary trajectory of combinatorial mutations in *E. coli*. For example, a six-site RBS library targeting membrane genes was found via TRACE to influence isobutanol tolerance.

3.7. Insights on Natural Evolution Learned from Directed Evolution

Directed evolution has been used as a tool to investigate molecular evolution mechanisms and obtain new strategies and guidance for performing efficient laboratory evolutionary experiments.[601,602] To gain insights on how evolution influences protein stability and vice versa, Bershtein and co-workers performed directed evolution on the *TEM-1* β-lactamase. They investigated how deleterious mutations and epistatic effects changed the protein's fitness landscape, subjecting to random mutation drift and purifying to produce changes in the fitness landscape. Using both computational simulations and experimental analyses, they found that robustness and epistasis are strongly correlated. Any protein system with threshold robustness is inevitably epistatic, meaning that if the stability threshold is exceeded, the protein will become far less robust.[603] The evolvability of proteins has two contradicting elements: robustness and innovability, where robustness refers to the ability of a protein to stay stable upon mutation; and innovability refers to the introduction of new function with mutations.[604] To decipher the relationship between these two conflicting characteristics, Dellus-Gur and co-workers performed directed evolution on *TEM-1* β-lactamase to obtain highly stabilized variants of *TEM-1* and showed that structural measures of fold polarity are correlated with protein innovability. Upon investigating these variants with new functions, they argued that highly functional variants exhibit selective rigidification of the structural scaffold while maintaining the conformation.[605]

Divergent and convergent evolution are two standard processes of natural evolution. The former is when groups from the same ancestor evolve to accumulate differences, finally forming new species. By contrast, in the latter process, different organisms independently evolve to adapt to the same environment, resulting in similar phenotypic outcomes but different genetic changes. Directed evolution can be used to investigate evolutionary divergence and convergence at the single protein level. For example, divergent evolutionary trajectories that were hard to reconstruct in natural systems were mimicked using the *de novo* designed bifunctional protein Syn-IF.[606] Syn-IF is used to rescue two auxotrophic *E. coli* variants: *ilvA*Δ and *fes*Δ. *IlvA* and *Fes* catalyze isoleucine biosynthesis and iron assimilation, respectively. Directed evolution of Syn-IF for fast rescued either Δ*ilvA* or Δ*fes* resulted in two monofunctional proteins that can rescue the selected function with enhanced activity over the parent one, but they lost the ability to rescue the unselected function.[606] In another study, a *de novo* zinc-binding artificial peptide was evolved to become a globular enzyme for easter cleavage. Directed evolution converted the zinc ion from an essential structural element into a catalytic cofactor.[607] Convergency is relatively easy to mimic by evolving different cell populations under similar selection pressure.[608] However, direct examination of convergency at the single-molecule level without interference from the genomic background is challenging. Liu and co-workers used PACE to directly investigate protein evolutionary convergence and reproducibility in a continuous format.[322] In this study, two independent T7 RNA polymerases were first evolved divergently to recognize two distinct intermediate promoters, and then evolved convergently to the same promoter but with different phenotypic activity. The different genotypes between the evolved populations and specific epistatic interactions showed that evolution is pathway-dependent and irreproducible; additionally, stochasticity among populations under the same environment determined the final fitness optimum each population could achieve.

Cryptic genetic variation plays an important role in adaptive evolution. Phenotype has a certain tolerance to genetic changes, resulting in several cryptic genetic variations that do not have effects on phenotype under specific genetic and environmental backgrounds but affect an organism's evolutionary behavior after genetic perturbations and environmental changes.[609−612] To elucidate how cryptic genetic variation affects adaptive evolution, Wagner and co-workers created cryptic variation by randomly mutating a well-characterized ribozyme while preserving its native function of RNA phosphate bond cleavage. The resulting variants adapted more rapidly to cleave RNA molecules containing a phosphorothioate linkage than the population without cryptic variation. Detailed sequence analysis revealed that cryptic variations expanded the population's genotypes in sequence space, and that some of these were preadapted to a changing environment.[613] Recently, they continued this study by creating four parallel populations of yellow fluorescent proteins (V^C) with cryptic variations using directed evolution in *E. coli*. Compared to those variants evolved from the wild-type (V^0), V^C adapted about three times faster to produce variants with green fluorescence, among which three showed significantly greater green fluorescence. The study of genotypes and accessibility of mutational paths revealed that V^C evolved adaptive genotypes with greater diversity and higher fitness than V^0. The neutral or deleterious mutations in V^C served as stepping-stones, facilitating fitness valley crossing and further evolving toward the new phenotype that cannot be accessed by starting from V^0.[614] The neutral drift strategy is similar to creating cryptic genetic variants and has been widely used in the directed evolution of proteins to access new reactivity.[615,616]

Additivity or nonadditivity arises when more than one mutation are introduced in a protein sequence, which plays an important role in protein engineering. Early directed evolution studies provided many examples of mutational additivity and it is common to find that mutated residues are not in molecular contact or are linked with different unrelated properties.[617−619] Nonadditivity, on the other hand, is less common to be addressed in the directed evolution study. Reetz and co-workers elucidated the mutational nonadditivity that contributes to the efficiency of a given mutagenesis strategy.[620] In one of their studies, deconvolution of an engineered highly stereoselective lipase PALmut (M16A, L17F, L162D) showed that L162D and M16A/L17F interacted in a strongly cooperative nonadditive manner since both mutants with L162D or M16A/L17F alone had very poor selectivity.[621]

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Another comprehensive deconvolution study was conducted on five sets of mutations obtained through ISM of the epoxide hydrolase from *Aspergillus niger*. The free energy of interaction $\Delta G^{\#}(i,j)$ between any two sets of mutations were calculated to prove that nonadditivity, both cooperative and antagonistic, are very common among different mutation pathways.[622] These studies uncovered that nonadditivity plays a crucial role in the efficacy of ISM strategy.

4. CONCLUSION AND FUTURE PERSPECTIVES

Directed evolution has been successfully used to evolve proteins with improved or novel functions *in vitro*. The availability of comprehensive genetic diversification methods enables the creation of a very diverse library of variants. However, limited by the throughput of both library creation and library screening/selection, only a small fraction of the entire protein sequence space can be explored. The proper starting point and mutation targets play key roles in successfully obtaining the desired function within a practical time scale in directed evolution experiments. The increased pairing of rational design and directed evolution can potentially allow researchers to predict beforehand the mutations that would confer a desired function. Sophisticated knowledge of protein structure, function, and dynamics will undoubtedly assist the design of directed evolution experiments. Additionally, recent advances in the development of sequence-based, structure-based, MD-based, and machine learning-based computational tools will facilitate the identification of the beneficial mutations and accelerate the protein engineering process by creating smaller but smarter libraries.[623]

Of particular note, machine learning can be seamlessly combined with directed evolution and we expect there will be many new advances along this frontier. Benefited by the rapid advances in DNA sequencing technologies, the amount of raw amino acid sequence information is rising exponentially. However, most of these sequencing data are unlabeled. Current machine learning models are largely supervised, which means that they are not able to utilize such unlabeled sequencing data. However, the information embedded in the unlabeled data is indeed helpful to applications such as protein property prediction.[177] To take full advantage of unlabeled sequence data, machine learning models trained in an unsupervised manner are promising. Some attempts, such as Unirep, have been proven effective and accurate for protein engineering tasks when combined with a supervised top model.[177] An alternative model, the transformer model, is widely applied as a language model in natural language processing tasks.[624] Due to the similarity between amino acid sequences and natural language, the transformer model can be used to embed protein sequences.[165] However, there is still potential for improving accurate embedding of amino acid sequences using uncharacterized sequences, as shown in the benchmark studies.[174] Another significant aspect that is rarely discussed is how to design the training data to maximize prediction power. A recent massive parallel study demonstrated that training data with varying sizes and diversity resulted in different model performance and highlighted the value of optimizing training data sets.[625] Thus, the methodology for designing training data for machine learning-assisted directed evolution remains for further investigation.

As emphasized several times, one of the biggest obstacles in directed evolution is the development of a proper screening/selection method. Thanks to tremendous efforts of scientists in the past decades, fast development of screening/selection methods has become possible. As a result, the chances of obtaining a variant with the desired properties are greatly enhanced. However, current screening/selection methods still have limitations that need to be addressed. Display-based selections are mostly limited to selecting binding proteins, while *in vivo* and *in vitro* compartmentalization methods are not suitable for phenotyping enzymes with conditions that are incompatible for both transcription-translation and selection.[281] On the other hand, some screening methods such as microtiter plates suffer from laborious, time-consuming sample preparation and generally low throughput. Although FACS shows extremely high sensitivity and throughput, the target enzyme must be coupled with a fluorescent protein for phenotyping. Ultrahigh-throughput droplet microfluidics improves the efficiency of directed evolution and enables the potential of finding rare targets among a large library. However, most droplet sorters use fluorescence or absorbance signals as the readout, which has limitations.

To tackle this key limitation, development of new technologies and the interplay between different advanced techniques such as automation and MS-based screening will be necessary to further improve assay creation, sample preparation, variant measurements, and data analysis. Automation can be applied at different stages in directed evolution to reduce cost and time, thus improving throughput. Robotic liquid handling approaches based on pipet tips,[626,627] pins,[628] inkjet, and acoustic deposition[629,630] have been applied to accelerate assay setup and sample preparation. Moreover, during screening/selection, automated introduction of samples into a mass spectrometer enables substantial improvements in analytic throughput. A single-probe autosampler is now a standard component for most GC-MS and LC-MS instrumentation for automatic flow injection, and multiprobe autosamplers have also been developed.[631] The RapidFire system[632,633] automates sample aspiration, solid-phase-extraction, and ESI-MS injection steps to achieve a cycling time of ~10 s. The NanoMate system[634,635] automates direct sample infusion from a 96-well plate to a chip-based nanoESI source. Besides automation, several sample preparation techniques and mass spectrometers can be coupled together to increase throughput, accuracy, and/or resolution of an analysis during directed evolution. An ESI source can be linked with LC,[636] CE,[637] direct infusion,[638,639] or desorption/ionization approaches to improve efficiency. Tandem MS[640] and triple quadrupole (QQQ)[641] MS are generally used for targets fragmentation and quantitation, respectively. Ion cyclotron resonance (ICR)[642] and orbitrap[643] mass analyzers are known for their superior mass accuracy with a resolution above 100,000 dots per inch. All these features can be potentially used for directed evolution to increase the chances of finding the desired target. However, special, yet expensive instrumentation, well-trained technicians, and bioinformatic pipelines are required to set up MS-based protein assays and interpret complex data sets. Typically, baseline correction, noise reduction, spectral alignment, and normalization are implemented to reduce variation from sample preparation, technical instrumental errors, or random human errors.[644] User friendly web-front interfaces such as microMS[645] or OpenMS[646] can be potentially used for quick preliminary analysis of data obtained from rounds of directed evolution.

Similarly, continuous evolution can overcome previous key limitations in directed evolution. Continuous evolution has

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

advantages over traditional directed evolution strategies because it operates without human intervention and can uncover more complex fitness landscapes. Although many targeted *in vivo* gene diversification methods have been developed in both prokaryotes and eukaryotes, the application of continuous evolution methods for improving phenotypes unrelated to cell growth is still rare. Moreover, it remains an overwhelming challenge to adapt existing continuous evolution methods for new target proteins, pathways, and genomes. One potential strategy to address this limitation is to integrate traditional directed evolution with automation and machine learning.[647] Automation can expedite directed evolution and provide data sets at a capacity that cannot be achieved manually.[591] However, automation alone is not enough to realize continuous evolution. The data generated from automated experiments can train and refine machine learning models that can, in turn, automatically predict beneficial mutations and guide the design of evolution experiments. In principle, the resulting fully closed "design-build-test-learn" cycle would enable directed evolution of any gene, pathway, or genome in a continuous manner.

In conclusion, directed evolution has been widely used to engineer nucleic acids, proteins, genetic circuits, biochemical pathways, and whole genomes with new or improved functions for basic and applied biological and medical research. With recent advances in sophisticated *in vivo* gene diversification methods, comprehensive high-throughput screening and automation platforms, and advanced computational design and machine learning tools, directed evolution is entering a new era to create more possibilities by revealing the secrets of natural evolution, biomolecular structure−function relationships, genotype-phenotype relationships, and engineering useful biological systems with a plethora of applications in the energy, chemical, pharmaceutical, agriculture, and food industries.

AUTHOR INFORMATION

Corresponding Author

Huimin Zhao − *Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, DOE Center for Advanced Bioenergy and Bioproducts Innovation, and Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-9069-6739; Phone: (217) 333-2631; Email: zhao5@illinois.edu; Fax: (217) 333-5052

Authors

Yajie Wang − *Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Pu Xue − *Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0001-6023-1298

Mingfeng Cao − *DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-6750-3871

Tianhao Yu − *Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Stephan T. Lane − *Carl R. Woese Institute for Genomic Biology and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrev.1c00260

Author Contributions

#Y.W. and P.X. contributed equally.

Notes

The authors declare no competing financial interest.

Biographies

Yajie Wang received her Ph.D. degree in Chemical and Biomolecular Engineering from the University of Illinois at Urbana−Champaign (UIUC) in 2018. She is now a postdoctoral fellow in the Department of Energy Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) at UIUC. Her research focuses on developing biocatalysis and chemoenzymatic catalysis systems for selective synthesis of bioactive compounds. In addition, she is developing *in vivo* continuous evolution strategies for protein and pathway engineering in yeast.

Pu (Mason) Xue graduated from the Rensselaer Polytechnic Institute (RPI) with a BS degree in Chemical and Biomolecular Engineering and joined the graduate program in Chemical and Biomolecular Engineering at UIUC in 2016. His Ph.D. thesis research focuses on developing a fully automated workflow to speed up the design−build−test−learn cycle for engineering of microbial cell factories. He has applied multiple synthetic biology and directed evolution concepts/tools for high-throughput library construction and screening during his Ph.D. thesis research.

Mingfeng Cao received his Ph.D. degree in Microbiology from the Nankai University in 2011. He is now a research scientist in CABBI at UIUC after working as a research associate at the Tianjin Institute of Industrial Biotechnology, Chinese Academy of Science (2011−2013) and as a postdoctoral research associate at the Iowa State University (2013−2018). His research focuses on the genetic tool development and metabolic engineering of nonconventional yeasts to produce value-added chemicals.

Tianhao Yu received his B.S. degree in Chemical Engineering from the University of Rochester in 2019. He is currently pursuing a Ph.D. degree from UIUC with Dr. Huimin Zhao. His research focuses on developing machine learning strategies for protein engineering.

Stephan Lane graduated from the UIUC with a BS degree in Molecular and Cellular Biology followed by a Ph.D. in Food Science and Human Nutrition under the mentorship of Dr. Yong-Su Jin, where his thesis research focused on metabolic engineering of yeasts using rational and evolutionary techniques. After graduation, he joined the lab of Dr. Huimin Zhao as manager of the Illinois Biofoundry for Advanced Biomanufacturing (iBioFAB), aiming to automate the design−build−test−learn cycle and develop new techniques in automated biological research.

Dr. Huimin Zhao is the Steven L. Miller Chair of Chemical and Biomolecular Engineering and Professor of Chemistry, Biochemistry, and Bioengineering at UIUC. He received his B.S. degree in Biology

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

from the University of Science and Technology of China in 1992 and his Ph.D. degree in Chemistry from the California Institute of Technology in 1998 under the guidance of Dr. Frances Arnold. He joined UIUC in 2000 after a stint as a project leader at The Dow Chemical Company and was promoted to full professor in 2008. Dr. Zhao has authored and coauthored over 360 research articles and over 30 issued and pending patent applications. In addition, he has given over 410 plenary, keynote, or invited lectures. His primary research interests are in the development and applications of directed evolution, synthetic biology, machine learning, and laboratory automation tools to address society's most daunting challenges in health, energy, and sustainability, and in the fundamental aspects of enzyme catalysis, cell metabolism, and gene regulation.

ACKNOWLEDGMENTS

We acknowledge financial support from the U.S. Department of Energy (DE-SC0018420 and DE-SC0018260) and the National Science Foundation (Award No. 2019897) for various directed evolution projects (H.Z.). The online tool of BioRender (BioRender.com) was used to create parts of Figures 6, 8−16, and 22.

REFERENCES

(1) Kacian, D. L.; Mills, D. R.; Kramer, F. R.; Spiegelman, S. A Replicating RNA Molecule Suitable for a Detailed Analysis of Extracellular Evolution and Replication. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 3038−3042.
(2) Levisohn, R.; Spiegelman, S. Further Extracellular Darwinian Experiments with Replicating RNA Molecules: Diverse Variants Isolated under Different Selective Conditions. *Proc. Natl. Acad. Sci. U. S. A.* **1969**, *63*, 805−811.
(3) Mills, D. R.; Peterson, R. L.; Spiegelman, S. An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 217−224.
(4) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228*, 1315−1317.
(5) Eigen, M.; Gardiner, W. Evolutionary Molecular Engineering Based on RNA Replication. *Pure Appl. Chem.* **1984**, *56*, 967−978.
(6) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818−822.
(7) Liao, H.; McKenzie, T.; Hageman, R. Isolation of a Thermostable Enzyme Variant by Cloning and Selection in a Thermophile. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 576−580.
(8) Zhao, H.; Chockalingam, K.; Chen, Z. Directed Evolution of Enzymes and Pathways for Industrial Biocatalysis. *Curr. Opin. Biotechnol.* **2002**, *13*, 104−110.
(9) Arnold, F. H.; Volkov, A. A. Directed Evolution of Biocatalysts. *Curr. Opin. Chem. Biol.* **1999**, *3*, 54−59.
(10) Arnold, F. H. Combinatorial and Computational Challenges for Biocatalyst Design. *Nature* **2001**, *409*, 253−257.
(11) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem., Int. Ed.* **2015**, *54*, 3351−3367.
(12) Hyster, T. K.; Ward, T. R. Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chem., Int. Ed.* **2016**, *55*, 7344−7357.
(13) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem., Int. Ed.* **2018**, *57*, 4143−4148.
(14) Reetz, M. T.; Zonta, A.; Schimossek, K.; Jaeger, K.-E.; Liebeton, K. Creation of Enantioselective Biocatalysts for Organic Chemistry by in Vitro Evolution. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2830−2832.
(15) Wang, Y.; Yu, X.; Zhao, H. Biosystems Design by Directed Evolution. *AIChE J.* **2020**, *66*, No. e16716.
(16) Chen, K.; Arnold, F. H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, *3*, 203−213.
(17) Tan, Z. L.; Zheng, X.; Wu, Y.; Jian, X.; Xing, X.; Zhang, C. In Vivo Continuous Evolution of Metabolic Pathways for Chemical Production. *Microb. Cell Fact.* **2019**, *18*, 82−100.
(18) Chica, R. A.; Doucet, N.; Pelletier, J. N. Semi-Rational Approaches to Engineering Enzyme Activity: Combining the Benefits of Directed Evolution and Rational Design. *Curr. Opin. Biotechnol.* **2005**, *16*, 378−384.
(19) Lutz, S.; Patrick, W. M. Novel Methods for Directed Evolution of Enzymes: Quality, Not Quantity. *Curr. Opin. Biotechnol.* **2004**, *15*, 291−297.
(20) Leung, D. W.; Chen, E.; Goeddel, D. A Method for Random Mutagenesis of a Defined DNA Segment Using a Modified Polymerase Chain Reaction. *Technique* **1989**, *1*, 11−15.
(21) 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.* **1996**, *255*, 589−603.
(22) Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. Reducing Mutational Bias in Random Protein Libraries. *Anal. Biochem.* **2005**, *339*, 9−14.
(23) Fujii, R.; Kitaoka, M.; Hayashi, K. One-Step Random Mutagenesis by Error-Prone Rolling Circle Amplification. *Nucleic Acids Res.* **2004**, *32*, No. e145.
(24) Fire, A.; Xu, S. Q. Rolling Replication of Short DNA Circles. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 4641−4645.
(25) Fujii, R.; Kitaoka, M.; Hayashi, K. Error-Prone Rolling Circle Amplification: The Simplest Random Mutagenesis Protocol. *Nat. Protoc.* **2006**, *1*, 2493−2497.
(26) Wong, T. S.; Tee, K. L.; Hauer, B.; Schwaneberg, U. Sequence Saturation Mutagenesis (SeSaM): A Novel Method for Directed Evolution. *Nucleic Acids Res.* **2004**, *32*, No. 26e.
(27) Wong, T. S.; Roccatano, D.; Loakes, D.; Tee, K. L.; Schenk, A.; Hauer, B.; Schwaneberg, U. Transversion-Enriched Sequence Saturation Mutagenesis (SeSaM-Tv+): A Random Mutagenesis Method with Consecutive Nucleotide Exchanges That Complements the Bias of Error-Prone PCR. *Biotechnol. J.* **2008**, *3*, 74−82.
(28) Mundhada, H.; Marienhagen, J.; Scacioc, A.; Schenk, A.; Roccatano, D.; Schwaneberg, U. SeSaM-Tv-II Generates a Protein Sequence Space That Is Unobtainable by epPCR. *ChemBioChem* **2011**, *12*, 1595−1601.
(29) Ruff, A. J.; Marienhagen, J.; Verma, R.; Roccatano, D.; Genieser, H.-G.; Niemann, P.; Shivange, A. V.; Schwaneberg, U. dRTP and dPTP a Complementary Nucleotide Couple for the Sequence Saturation Mutagenesis (SeSaM) Method. *J. Mol. Catal. B: Enzym.* **2012**, *84*, 40−47.
(30) Murakami, H.; Hohsaka, T.; Sisido, M. Random Insertion and Deletion of Arbitrary Number of Bases for Codon-Based Random Mutation of DNAs. *Nat. Biotechnol.* **2002**, *20*, 76−81.
(31) Haapa, S.; Taira, S.; Heikkinen, E.; Savilahti, H. An Efficient and Accurate Integration of Mini-Mu Transposons in Vitro: A General Methodology for Functional Genetic Analysis and Molecular Biology Applications. *Nucleic Acids Res.* **1999**, *27*, 2777−2784.
(32) Matteucci, M. D.; Heyneker, H. L. Targeted Random Mutagenesis: The Use of Ambiguously Synthesized Oligonucleotides to Mutagenize Sequences Immediately 5′ of an ATG Initiation Codon. *Nucleic Acids Res.* **1983**, *11*, 3113−3121.
(33) Reidhaar-Olson, J. F.; Sauer, R. T. Combinatorial Cassette Mutagenesis as a Probe of the Informational Content of Protein Sequences. *Science* **1988**, *241*, 53−57.
(34) Wells, J. A.; Vasser, M.; Powers, D. B. Cassette Mutagenesis: An Efficient Method for Generation of Multiple Mutations at Defined Sites. *Gene* **1985**, *34*, 315−323.
(35) Byrappa, S.; Gavin, D. K.; Gupta, K. C. A Highly Efficient Procedure for Site-Specific Mutagenesis of Full-Length Plasmids Using Vent DNA Polymerase. *Genome Res.* **1995**, *5*, 404−407.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(36) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**, *8*, 91−100.
(37) Tseng, W.-C.; Lin, J.-W.; Wei, T.-Y.; Fang, T.-Y. A Novel Megaprimed and Ligase-Free, PCR-Based, Site-Directed Mutagenesis Method. *Anal. Biochem.* **2008**, *375*, 376−378.
(38) Dennig, A.; Shivange, A. V.; Marienhagen, J.; Schwaneberg, U. OmniChange: The Sequence Independent Method for Simultaneous Site-Saturation of Five Codons. *PLoS One* **2011**, *6*, No. e26222.
(39) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343−345.
(40) Püllmann, P.; Ulpinnis, C.; Marillonnet, S.; Gruetzner, R.; Neumann, S.; Weissenborn, M. J. Golden Mutagenesis: An Efficient Multi-Site-Saturation Mutagenesis Approach by Golden Gate Cloning with Automated Primer Design. *Sci. Rep.* **2019**, *9*, 10932−10942.
(41) Quan, J.; Tian, J. Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *PLoS One* **2009**, *4*, No. e6441.
(42) An, Y.; Ji, J.; Wu, W.; Lv, A.; Huang, R.; Wei, Y. A Rapid and Efficient Method for Multiple-Site Mutagenesis with a Modified Overlap Extension PCR. *Appl. Microbiol. Biotechnol.* **2005**, *68*, 774−778.
(43) Peng, R.-H.; Xiong, A.-S.; Yao, Q.-H. A Direct and Efficient Page-Mediated Overlap Extension PCR Method for Gene Multiple-Site Mutagenesis. *Appl. Microbiol. Biotechnol.* **2006**, *73*, 234−240.
(44) Seyfang, A.; Jin, J. H. Multiple Site-Directed Mutagenesis of More Than 10 Sites Simultaneously and in a Single Round. *Anal. Biochem.* **2004**, *324*, 285−291.
(45) Herman, A.; Tawfik, D. S. Incorporating Synthetic Oligonucleotides Via Gene Reassembly (ISOR): A Versatile Tool for Generating Targeted Libraries. *Protein Eng. Des. Sel.* **2007**, *20*, 219−226.
(46) Reetz, M. T.; Wu, S. Greatly Reduced Amino Acid Alphabets in Directed Evolution: Making the Right Choice for Saturation Mutagenesis at Homologous Enzyme Positions. *Chem. Commun.* **2008**, 5499−5501.
(47) Jochens, H.; Bornscheuer, U. T. Natural Diversity to Guide Focused Directed Evolution. *ChemBioChem* **2010**, *11*, 1861−1866.
(48) Xu, J.; Cen, Y.; Singh, W.; Fan, J.; Wu, L.; Lin, X.; Zhou, J.; Huang, M.; Reetz, M. T.; Wu, Q. Stereodivergent Protein Engineering of a Lipase to Access All Possible Stereoisomers of Chiral Esters with Two Stereocenters. *J. Am. Chem. Soc.* **2019**, *141*, 7934−7945.
(49) Reetz, M. T.; Carballeira, J. D. Iterative Saturation Mutagenesis (ISM) for Rapid Directed Evolution of Functional Enzymes. *Nat. Protoc.* **2007**, *2*, 891−903.
(50) Chockalingam, K.; Chen, Z.; Katzenellenbogen, J. A.; Zhao, H. Directed Evolution of Specific Receptor−Ligand Pairs for Use in the Creation of Gene Switches. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5691−5696.
(51) Reetz, M. T.; Wang, L.-W.; Bocola, M. Directed Evolution of Enantioselective Enzymes: Iterative Cycles of Casting for Probing Protein-Sequence Space. *Angew. Chem., Int. Ed.* **2006**, *45*, 1236−1241.
(52) Reetz, M. T.; Carballeira, J. D.; Vogel, A. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angew. Chem., Int. Ed.* **2006**, *45*, 7745−7751.
(53) Holland, J. H. Genetic Algorithms. *Sci. Am.* **1992**, *267*, 66−73.
(54) Holland, J. H. *Adaptation in Natural and Artificial Systems*; MIT Press: Cambridge, 1992.
(55) Stemmer, W. P. C. Rapid Evolution of a Protein in Vitro by DNA Shuffling. *Nature* **1994**, *370*, 389−391.
(56) Zhang, J.-H.; Dawes, G.; Stemmer, W. P. C. Directed Evolution of a Fucosidase from a Galactosidase by DNA Shuffling and Screening. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4504−4509.
(57) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat. Biotechnol.* **1996**, *14*, 315−319.
(58) Crameri, A.; Dawes, G.; Rodriguez, E., Jr; Silver, S.; Stemmer, W. P. Molecular Evolution of an Arsenate Detoxification Pathway by DNA Shuffling. *Nat. Biotechnol.* **1997**, *15*, 436−438.
(59) Crameri, A.; Raillard, S. A.; Bermudez, E.; Stemmer, W. P. DNA Shuffling of a Family of Genes from Diverse Species Accelerates Directed Evolution. *Nature* **1998**, *391*, 288−291.
(60) Ness, J. E.; Welch, M.; Giver, L.; Bueno, M.; Cherry, J. R.; Borchert, T. V.; Stemmer, W. P. C.; Minshull, J. DNA Shuffling of Subgenomic Sequences of Subtilisin. *Nat. Biotechnol.* **1999**, *17*, 893−896.
(61) Joo, H.; Lin, Z.; Arnold, F. H. Laboratory Evolution of Peroxide-Mediated Cytochrome P450 Hydroxylation. *Nature* **1999**, *399*, 670−673.
(62) Cherry, J. R.; Fidantsef, A. L. Directed Evolution of Industrial Enzymes: An Update. *Curr. Opin. Biotechnol.* **2003**, *14*, 438−443.
(63) Giver, L.; Gershenson, A.; Freskgard, P.-O.; Arnold, F. H. Directed Evolution of a Thermostable Esterase. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12809−12813.
(64) Zhao, H.; Arnold, F. H. Directed Evolution Converts Subtilisin E into a Functional Equivalent of Thermitase. *Protein Eng., Des. Sel.* **1999**, *12*, 47−53.
(65) Kikuchi, M.; Ohnishi, K.; Harayama, S. Novel Family Shuffling Methods for the in Vitro Evolution of Enzymes. *Gene* **1999**, *236*, 159−167.
(66) Buchholz, F.; Angrand, P.-O.; Stewart, A. F. Improved Properties of FLP Recombinase Evolved by Cycling Mutagenesis. *Nat. Biotechnol.* **1998**, *16*, 657−662.
(67) Ness, J. E.; Kim, S.; Gottman, A.; Pak, R.; Krebber, A.; Borchert, T. V.; Govindarajan, S.; Mundorff, E. C.; Minshull, J. Synthetic Shuffling Expands Functional Protein Diversity by Allowing Amino Acids to Recombine Independently. *Nat. Biotechnol.* **2002**, *20*, 1251−1255.
(68) Stemmer, W. P. DNA Shuffling by Random Fragmentation and Reassembly: In Vitro Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10747−10751.
(69) Stemmer, W. P.; Crameri, 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* **1995**, *164*, 49−53.
(70) Joern, J. M.; Meinhold, P.; Arnold, F. H. Analysis of Shuffled Gene Libraries. *J. Mol. Biol.* **2002**, *316*, 643−656.
(71) Coco, W. M.; Levinson, W. E.; Crist, M. J.; Hektor, H. J.; Darzins, A.; Pienkos, P. T.; Squires, C. H.; Monticello, D. J. DNA Shuffling Method for Generating Highly Recombined Genes and Evolved Enzymes. *Nat. Biotechnol.* **2001**, *19*, 354−359.
(72) Abécassis, V.; Pompon, D.; Truan, G. High Efficiency Family Shuffling Based on Multi-Step PCR and in Vivo DNA Recombination in Yeast: Statistical and Functional Analysis of a Combinatorial Library between Human Cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **2000**, *28*, No. 88e.
(73) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. Molecular Evolution by Staggered Extension Process (StEP) in Vitro Recombination. *Nat. Biotechnol.* **1998**, *16*, 258−261.
(74) Mitra, B.; Gerlt, J. A.; Babbitt, P. C.; Koo, C. W.; Kenyon, G. L.; Joseph, D.; Petsko, G. A. A Novel Structural Basis for Membrane Association of a Protein: Construction of a Chimeric Soluble Mutant of (S)-Mandelate Dehydrogenase from Pseudomonas Putida. *Biochemistry* **1993**, *32*, 12959−12967.
(75) Bogarad, L. D.; Deem, M. W. A Hierarchical Approach to Protein Molecular Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 2591−2595.
(76) Ostermeier, M.; Shim, J. H.; Benkovic, S. J. A Combinatorial Approach to Hybrid Enzymes Independent of DNA Homology. *Nat. Biotechnol.* **1999**, *17*, 1205−1209.
(77) Lutz, S.; Ostermeier, M.; Moore, G. L.; Maranas, C. D.; Benkovic, S. J. Creating Multiple-Crossover DNA Libraries Independent of Sequence Identity. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 11248−11253.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(78) Sieber, V.; Martinez, C. A.; Arnold, F. H. Libraries of Hybrid Proteins from Distantly Related Sequences. *Nat. Biotechnol.* **2001**, *19*, 456−460.
(79) 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. U. S. A.* **2004**, *101*, 7011−7016.
(80) Hiraga, K.; Arnold, F. H. General Method for Sequence-Independent Site-Directed Chimeragenesis. *J. Mol. Biol.* **2003**, *330*, 287−296.
(81) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS One* **2009**, *4*, No. e5553.
(82) Voigt, C. A.; Martinez, C.; Wang, Z.-G.; Mayo, S. L.; Arnold, F. H. Protein Building Blocks Preserved by Recombination. *Nat. Struct. Biol.* **2002**, *9*, 553−558.
(83) Cox, E. C. Bacterial Mutator Genes and the Control of Spontaneous Mutation. *Annu. Rev. Genet.* **1976**, *10*, 135−156.
(84) Schaaper, R. M. Mechanisms of Mutagenesis in the Escherichia coli Mutator mutD5: Role of DNA Mismatch Repair. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 8126−8130.
(85) Scheuermann, R.; Tam, S.; Burgers, P. M.; Lu, C.; Echols, H. Identification of the Epsilon-Subunit of Escherichia coli DNA Polymerase III Holoenzyme as the dnaQ Gene Product: A Fidelity Subunit for DNA Replication. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 7085−7089.
(86) Greener, A.; Callahan, M.; Jerpseth, B. An Efficient Random Mutagenesis Technique Using an *E. coli* Mutator Strain. *Mol. Biotechnol.* **1997**, *7*, 189−195.
(87) Selifonova, O.; Valle, F.; Schellenberger, V. Rapid Evolution of Novel Traits in Microorganisms. *Appl. Environ. Microbiol.* **2001**, *67*, 3645−3649.
(88) Zhang, Y. X.; Perry, K.; Vinci, V. A.; Powell, K.; Stemmer, W. P.; del Cardayre, S. B. Genome Shuffling Leads to Rapid Phenotypic Improvement in Bacteria. *Nature* **2002**, *415*, 644−646.
(89) Patnaik, R.; Louie, S.; Gavrilovic, V.; Perry, K.; Stemmer, W. P.; Ryan, C. M.; del Cardayre, S. Genome Shuffling of Lactobacillus for Improved Acid Tolerance. *Nat. Biotechnol.* **2002**, *20*, 707−712.
(90) Kao, K. N.; Constabel, F.; Michayluk, M. R.; Gamborg, O. L. Plant Protoplast Fusion and Growth of Intergeneric Hybrid Cells. *Planta* **1974**, *120*, 215−227.
(91) Gong, J.; Zheng, H.; Wu, Z.; Chen, T.; Zhao, X. Genome Shuffling: Progress and Applications for Phenotype Improvement. *Biotechnol. Adv.* **2009**, *27*, 996−1005.
(92) Chen, H.; Liu, S.; Padula, S.; Lesman, D.; Griswold, K.; Lin, A.; Zhao, T.; Marshall, J. L.; Chen, F. Efficient, Continuous Mutagenesis in Human Cells Using a Pseudo-Random DNA Editor. *Nat. Biotechnol.* **2020**, *38*, 165−168.
(93) Moore, C. L.; Papa, L. J., 3rd; Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions in Vivo. *J. Am. Chem. Soc.* **2018**, *140*, 11560−11564.
(94) Álvarez, B.; Mencía, M.; de Lorenzo, V.; Fernández, L. Á. In Vivo Diversification of Target Genomic Sites Using Processive Base Deaminase Fusions Blocked by dCas9. *Nat. Commun.* **2020**, *11*, 6436.
(95) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175*, 1946−1957.
(96) Rajewsky, K.; Forster, I.; Cumano, A. Evolutionary and Somatic Selection of the Antibody Repertoire in the Mouse. *Science* **1987**, *238*, 1088−1094.
(97) Papavasiliou, F. N.; Schatz, D. G. Somatic Hypermutation of Immunoglobulin Genes: Merging Mechanisms for Genetic Diversity. *Cell* **2002**, *109*, S35−S44.
(98) Cumbers, S. J.; Williams, G. T.; Davies, S. L.; Grenfell, R. L.; Takeda, S.; Batista, F. D.; Sale, J. E.; Neuberger, M. S. Generation and Iterative Affinity Maturation of Antibodies in Vitro Using Hypermutating B-Cell Lines. *Nat. Biotechnol.* **2002**, *20*, 1129−1134.
(99) Gordon, M. S.; Kanegai, C. M.; Doerr, J. R.; Wall, R. Somatic Hypermutation of the B Cell Receptor Genes. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4126−4131.
(100) Wang, C. L.; Harper, R. A.; Wabl, M. Genome-Wide Somatic Hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7352−7356.
(101) Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y. Evolution of New Nonantibody Proteins Via Iterative Somatic Hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16745−16749.
(102) Ellis, H. M.; Yu, D.; DiTizio, T.; Court, D. L. High Efficiency Mutagenesis, Repair, and Engineering of Chromosomal DNA Using Single-Stranded Oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 6742−6746.
(103) Costantino, N.; Court, D. L. Enhanced Levels of Lambda Red-Mediated Recombinants in Mismatch Repair Mutants. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15748−15753.
(104) Wang, H. H.; Isaacs, F. J.; Carr, P. A.; Sun, Z. Z.; Xu, G.; Forest, C. R.; Church, G. M. Programming Cells by Multiplex Genome Engineering and Accelerated Evolution. *Nature* **2009**, *460*, 894−898.
(105) Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; et al. Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **2011**, *333*, 348−353.
(106) Lennen, R. M.; Nilsson Wallin, A. I.; Pedersen, M.; Bonde, M.; Luo, H.; Herrgård, M. J.; Sommer, M. O. A. Transient Overexpression of DNA Adenine Methylase Enables Efficient and Mobile Genome Engineering with Reduced Off-Target Effects. *Nucleic Acids Res.* **2016**, *44*, No. e36-e36.
(107) Nyerges, Á.; Csörgő, B.; Draskovits, G.; Kintses, B.; Szili, P.; Ferenc, G.; Révész, T.; Ari, E.; Nagy, I.; Bálint, B.; et al. Directed Evolution of Multiple Genomic Loci Allows the Prediction of Antibiotic Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5726−E5735.
(108) Nyerges, A.; Csorgo, B.; Nagy, I.; Balint, B.; Bihari, P.; Lazar, V.; Apjok, G.; Umenhoffer, K.; Bogos, B.; Posfai, G.; et al. A Highly Precise and Portable Genome Engineering Method Allows Comparison of Mutational Effects across Bacterial Species. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2502−2507.
(109) Aronshtam, A.; Marinus, M. G. Dominant Negative Mutator Mutations in the MutI Gene of Escherichia coli. *Nucleic Acids Res.* **1996**, *24*, 2498−2504.
(110) Carr, P. A.; Wang, H. H.; Sterling, B.; Isaacs, F. J.; Lajoie, M. J.; Xu, G.; Church, G. M.; Jacobson, J. M. Enhanced Multiplex Genome Engineering through Co-Operative Oligonucleotide Co-Selection. *Nucleic Acids Res.* **2012**, *40*, No. e132.
(111) Wang, H. H.; Kim, H.; Cong, L.; Jeong, J.; Bang, D.; Church, G. M. Genome-Scale Promoter Engineering by Coselection MAGE. *Nat. Methods* **2012**, *9*, 591−593.
(112) Warner, J. R.; Reeder, P. J.; Karimpour-Fard, A.; Woodruff, L. B. A.; Gill, R. T. Rapid Profiling of a Microbial Genome Using Mixtures of Barcoded Oligonucleotides. *Nat. Biotechnol.* **2010**, *28*, 856−138.
(113) Storici, F.; Resnick, M. A. In *Genetic Engineering: Principles and Methods*; Setlow, J. K., Ed.; Springer US: Boston, MA, 2003.
(114) DiCarlo, J. E.; Conley, A. J.; Penttilä, M.; Jäntti, J.; Wang, H. H.; Church, G. M. Yeast Oligo-Mediated Genome Engineering (YOGE). *ACS Synth. Biol.* **2013**, *2*, 741−749.
(115) Barbieri, E. M.; Muir, P.; Akhuetie-Oni, B. O.; Yellman, C. M.; Isaacs, F. J. Precise Editing at DNA Replication Forks Enables Multiplex Genome Engineering in Eukaryotes. *Cell* **2017**, *171*, 1453−1467.
(116) Dymond, J.; Boeke, J. The Saccharomyces cerevisiae SCRaMbLE System and Genome Minimization. *Bioeng. Bugs* **2012**, *3*, 170−171.
(117) Dymond, J. S.; Richardson, S. M.; Coombes, C. E.; Babatz, T.; Muller, H.; Annaluru, N.; Blake, W. J.; Schwerzmann, J. W.; Dai, J.; Lindstrom, D. L.; et al. Synthetic Chromosome Arms Function in Yeast and Generate Phenotypic Diversity by Design. *Nature* **2011**, *477*, 471−476.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(118) Blount, B. A.; Gowers, G. F.; Ho, J. C. H.; Ledesma-Amaro, R.; Jovicevic, D.; McKiernan, R. M.; Xie, Z. X.; Li, B. Z.; Yuan, Y. J.; Ellis, T. Rapid Host Strain Improvement by in Vivo Rearrangement of a Synthetic Yeast Chromosome. *Nat. Commun.* **2018**, *9*, 1932.
(119) Jia, B.; Wu, Y.; Li, B.-Z.; Mitchell, L. A.; Liu, H.; Pan, S.; Wang, J.; Zhang, H.-R.; Jia, N.; Li, B.; et al. Precise Control of SCRaMbLE in Synthetic Haploid and Diploid Yeast. *Nat. Commun.* **2018**, *9*, 1933.
(120) Shen, M. J.; Wu, Y.; Yang, K.; Li, Y.; Xu, H.; Zhang, H.; Li, B. Z.; Li, X.; Xiao, W. H.; Zhou, X.; et al. Heterozygous Diploid and Interspecies SCRaMbLEing. *Nat. Commun.* **2018**, *9*, 1934−1941.
(121) Dymond, J. S.; Richardson, S. M.; Coombes, C. E.; Babatz, T.; Muller, H.; Annaluru, N.; Blake, W. J.; Schwerzmann, J. W.; Dai, J.; Lindstrom, D. L.; et al. Synthetic Chromosome Arms Function in Yeast and Generate Phenotypic Diversity by Design. *Nature* **2011**, *477*, 471−476.
(122) Hochrein, L.; Mitchell, L. A.; Schulz, K.; Messerschmidt, K.; Mueller-Roeber, B. L-SCRaMbLE as a Tool for Light-Controlled Cre-Mediated Recombination in Yeast. *Nat. Commun.* **2018**, *9*, 1931.
(123) Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nat. Commun.* **2018**, *9*, 1911.
(124) Anzalone, A. V.; Koblan, L. W.; Liu, D. R. Genome Editing with CRISPR−Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nat. Biotechnol.* **2020**, *38*, 824−844.
(125) Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A. RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems. *Nat. Biotechnol.* **2013**, *31*, 233−239.
(126) Bao, Z.; Xiao, H.; Liang, J.; Zhang, L.; Xiong, X.; Sun, N.; Si, T.; Zhao, H. Homology-Integrated CRISPR−Cas (HI-CRISPR) System for One-Step Multigene Disruption in Saccharomyces cerevisiae. *ACS Synth. Biol.* **2015**, *4*, 585−594.
(127) Ronda, C.; Pedersen, L. E.; Sommer, M. O.; Nielsen, A. T. CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* **2016**, *6*, 19452−19461.
(128) Li, Y.; Lin, Z.; Huang, C.; Zhang, Y.; Wang, Z.; Tang, Y.-j.; Chen, T.; Zhao, X. Metabolic Engineering of Escherichia coli Using CRISPR−Cas9 Meditated Genome Editing. *Metab. Eng.* **2015**, *31*, 13−21.
(129) Jakočiūnas, T.; Pedersen, L. E.; Lis, A. V.; Jensen, M. K.; Keasling, J. D. CasPER, a Method for Directed Evolution in Genomic Contexts Using Mutagenesis and CRISPR/Cas9. *Metab. Eng.* **2018**, *48*, 288−296.
(130) Bao, Z.; HamediRad, M.; Xue, P.; Xiao, H.; Tasan, I.; Chao, R.; Liang, J.; Zhao, H. Genome-Scale Engineering of Saccharomyces cerevisiae with Single-Nucleotide Precision. *Nat. Biotechnol.* **2018**, *36*, 505−508.
(131) Guo, X.; Chavez, A.; Tung, A.; Chan, Y.; Kaas, C.; Yin, Y.; Cecchi, R.; Garnier, S. L.; Kelsic, E. D.; Schubert, M.; et al. High-Throughput Creation and Functional Profiling of DNA Sequence Variant Libraries Using CRISPR−Cas9 in Yeast. *Nat. Biotechnol.* **2018**, *36*, 540−546.
(132) Roy, K. R.; Smith, J. D.; Vonesch, S. C.; Lin, G.; Tu, C. S.; Lederer, A. R.; Chu, A.; Suresh, S.; Nguyen, M.; Horecka, J.; et al. Multiplexed Precision Genome Editing with Trackable Genomic Barcodes in Yeast. *Nat. Biotechnol.* **2018**, *36*, 512−520.
(133) Garst, A. D.; Bassalo, M. C.; Pines, G.; Lynch, S. A.; Halweg-Edwards, A. L.; Liu, R.; Liang, L.; Wang, Z.; Zeitoun, R.; Alexander, W. G.; et al. Genome-Wide Mapping of Mutations at Single-Nucleotide Resolution for Protein, Metabolic and Genome Engineering. *Nat. Biotechnol.* **2017**, *35*, 48−55.
(134) Lian, J.; HamediRad, M.; Hu, S.; Zhao, H. Combinatorial Metabolic Engineering Using an Orthogonal Tri-Functional CRISPR System. *Nat. Commun.* **2017**, *8*, 1688.
(135) Mason, D. M.; Weber, C. R.; Parola, C.; Meng, S. M.; Greiff, V.; Kelton, W. J.; Reddy, S. T. High-Throughput Antibody Engineering in Mammalian Cells by CRISPR/Cas9-Mediated Homology-Directed Mutagenesis. *Nucleic Acids Res.* **2018**, *46*, 7436−7449.
(136) Erdogan, M.; Fabritius, A.; Basquin, J.; Griesbeck, O. Targeted in Situ Protein Diversification and Intra-Organelle Validation in Mammalian Cells. *Cell Chem. Biol.* **2020**, *27*, 610−621.
(137) Ma, Y.; Zhang, J.; Yin, W.; Zhang, Z.; Song, Y.; Chang, X. Targeted AID-Mediated Mutagenesis (TAM) Enables Efficient Genomic Diversification in Mammalian Cells. *Nat. Methods* **2016**, *13*, 1029−1035.
(138) Dominguez, A. A.; Lim, W. A.; Qi, L. S. Beyond Editing: Repurposing CRISPR-Cas9 for Precision Genome Regulation and Interrogation. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 5−15.
(139) Hess, G. T.; Frésard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C. Directed Evolution Using dCas9-Targeted Somatic Hypermutation in Mammalian Cells. *Nat. Methods* **2016**, *13*, 1036−1042.
(140) Li, C.; Zhang, R.; Meng, X.; Chen, S.; Zong, Y.; Lu, C.; Qiu, J. L.; Chen, Y. H.; Li, J.; Gao, C. Targeted, Random Mutagenesis of Plant Genes with Dual Cytosine and Adenine Base Editors. *Nat. Biotechnol.* **2020**, *38*, 875−882.
(141) Kuscu, C.; Parlak, M.; Tufan, T.; Yang, J.; Szlachta, K.; Wei, X.; Mammadov, R.; Adli, M. CRISPR-STOP: Gene Silencing through Base-Editing-Induced Nonsense Mutations. *Nat. Methods* **2017**, *14*, 710−712.
(142) Halperin, S. O.; Tou, C. J.; Wong, E. B.; Modavi, C.; Schaffer, D. V.; Dueber, J. E. CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a Tunable Window. *Nature* **2018**, *560*, 248−252.
(143) Lampson, B. C.; Inouye, M.; Inouye, S. Retrons, msDNA, and the Bacterial Genome. *Cytogenet. Genome Res.* **2005**, *110*, 491−499.
(144) Lim, D.; Maas, W. K. Reverse Transcriptase-Dependent Synthesis of a Covalently Linked, Branched DNA-RNA Compound in E. coli B. *Cell* **1989**, *56*, 891−904.
(145) Farzadfard, F.; Lu, T. K. Genomically Encoded Analog Memory with Precise in Vivo DNA Writing in Living Cell Populations. *Science* **2014**, *346*, 1256272.
(146) Simon, A. J.; Morrow, B. R.; Ellington, A. D. Retroelement-Based Genome Editing and Evolution. *ACS Synth. Biol.* **2018**, *7*, 2600−2611.
(147) Crook, N.; Abatemarco, J.; Sun, J.; Wagner, J. M.; Schmitz, A.; Alper, H. S. In Vivo Continuous Evolution of Genes and Pathways in Yeast. *Nat. Commun.* **2016**, *7*, 13051.
(148) Wilhelm, F. X.; Wilhelm, M.; Gabriel, A. Reverse Transcriptase and Integrase of the *Saccharomyces cerevisiae* Ty1 Element. *Cytogenet. Genome Res.* **2005**, *110*, 269−287.
(149) Gabriel, A.; Willems, M.; Mules, E. H.; Boeke, J. D. Replication Infidelity During a Single Cycle of Ty1 Retrotransposition. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 7767.
(150) Boutabout, M.; Wilhelm, M.; Wilhelm, F. X. DNA Synthesis Fidelity by the Reverse Transcriptase of the Yeast Retrotransposon Ty1. *Nucleic Acids Res.* **2001**, *29*, 2217−2222.
(151) Ng, P. C.; Henikoff, S. Sift: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Res.* **2003**, *31*, 3812−3814.
(152) Thomas, P. D.; Campbell, M. J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* **2003**, *13*, 2129−2141.
(153) Choi, Y.; Sims, G. E.; Murphy, S.; Miller, J. R.; Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* **2012**, *7*, e46688.
(154) Sebestova, E.; Bendl, J.; Brezovsky, J.; Damborsky, J. In *Directed Evolution Library Creation*; Springer, 2014.
(155) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues. *Nucleic Acids Res.* **2006**, *34*, W116−118.
(156) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; et al. CAVER 3.0: A Tool for the Analysis of Transport Pathways in

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, No. e1002708.
(157) Pavelka, A.; Chovancova, E.; Damborsky, J. Hotspot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering. *Nucleic Acids Res.* **2009**, *37*, W376−W383.
(158) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129−1143.
(159) Korendovych, I. V. Rational and Semirational Protein Design. *Methods Mol. Biol.* **2018**, *1685*, 15−23.
(160) Wang, S.; Fu, G.; Li, J.; Wei, X.; Fang, H.; Huang, D.; Lin, J.; Zhang, D. High-Efficiency Secretion and Directed Evolution of Chitinase BcChiA1 in Bacillus Subtilis for the Conversion of Chitinaceous Wastes into Chitooligosaccharides. *Front. Bioeng. Biotechnol.* **2020**, *8*, 432.
(161) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision Is Essential for Efficient Catalysis in an Evolved Kemp Eliminase. *Nature* **2013**, *503*, 418−421.
(162) Pavlova, M.; Klvana, M.; Prokop, Z.; Chaloupkova, R.; Banas, P.; Otyepka, M.; Wade, R. C.; Tsuda, M.; Nagata, Y.; Damborsky, J. Redesigning Dehalogenase Access Tunnels as a Strategy for Degrading an Anthropogenic Substrate. *Nat. Chem. Biol.* **2009**, *5*, 727−733.
(163) Jaeger, K.-E.; Eggert, T. Enantioselective Biocatalysis Optimized by Directed Evolution. *Curr. Opin. Biotechnol.* **2004**, *15*, 305−313.
(164) Ma, J.; Wu, L.; Guo, F.; Gu, J.; Tang, X.; Jiang, L.; Liu, J.; Zhou, J.; Yu, H. Enhanced Enantioselectivity of a Carboxyl Esterase from Rhodobacter Sphaeroides by Directed Evolution. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 4897−4906.
(165) Wu, S.; Acevedo, J. P.; Reetz, M. T. Induced Allostery in the Directed Evolution of an Enantioselective Baeyer−Villiger Monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2775−2780.
(166) Khersonsky, O.; Kiss, G.; Roethlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. Bridging the Gaps in Design Methodologies by Evolutionary Optimization of the Stability and Proficiency of Designed Kemp Eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10358−10363.
(167) Smith, S. N.; Wang, Y.; Baylon, J. L.; Singh, N. K.; Baker, B. M.; Tajkhorshid, E.; Kranz, D. M. Changing the Peptide Specificity of a Human T-Cell Receptor by Directed Evolution. *Nat. Commun.* **2014**, *5*, 5223.
(168) Liang, C.; Fioroni, M.; Rodríguez-Ropero, F.; Xue, Y.; Schwaneberg, U.; Ma, Y. Directed Evolution of a Thermophilic Endoglucanase (Cel5A) into Highly Active Cel5A Variants with an Expanded Temperature Profile. *J. Biotechnol.* **2011**, *154*, 46−53.
(169) Romero, P. A.; Krause, A.; Arnold, F. H. Navigating the Protein Fitness Landscape with Gaussian Processes. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E193−E201.
(170) Wu, Z.; Kan, S. B. J.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine Learning-Assisted Directed Protein Evolution with Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8852−8858.
(171) Yang, K. K.; Wu, Z.; Arnold, F. H. Machine-Learning-Guided Directed Evolution for Protein Engineering. *Nat. Methods* **2019**, *16*, 687−694.
(172) Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE. T. Pattern. Anal.* **2013**, *35*, 1798−1828.
(173) Bedbrook, C. N.; Yang, K. K.; Rice, A. J.; Gradinaru, V.; Arnold, F. H. Machine Learning to Design Integral Membrane Channelrhodopsins for Efficient Eukaryotic Expression and Plasma Membrane Localization. *PLoS Comput. Biol.* **2017**, *13*, No. e1005786.
(174) Rao, R.; Bhattacharya, N.; Thomas, N.; Duan, Y.; Chen, X.; Canny, J.; Abbeel, P.; Song, Y. S. Evaluating Protein Transfer Learning with TAPE. *Adv Neural Inf Process Syst.* **2019**, *32*, 9689−9701.
(175) Luo, Y.; Vo, L.; Ding, H.; Su, Y.; Liu, Y.; Qian, W. W.; Zhao, H.; Peng, J. Evolutionary Context-Integrated Deep Sequence Modeling for Protein Engineering. *bioRxiv* **2020**.
(176) Le, Q.; Mikolov, T. *International conference on machine learning*, 2014, 1188−1196.
(177) Alley, E. C.; Khimulya, G.; Biswas, S.; AlQuraishi, M.; Church, G. M. Unified Rational Protein Engineering with Sequence-Based Deep Representation Learning. *Nat. Methods* **2019**, *16*, 1315−1322.
(178) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805* **2018**.
(179) Li, Y.; Drummond, D. A.; Sawayama, A. M.; Snow, C. D.; Bloom, J. D.; Arnold, F. H. A Diverse Family of Thermostable Cytochrome P450s Created by Recombination of Stabilizing Fragments. *Nat. Biotechnol.* **2007**, *25*, 1051−1056.
(180) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; et al. Improving Catalytic Function by ProSAR-Driven Enzyme Evolution. *Nat. Biotechnol.* **2007**, *25*, 338−344.
(181) Biswas, S.; Khimulya, G.; Alley, E. C.; Esvelt, K. M.; Church, G. M. Low-N Protein Engineering with Data-Efficient Deep Learning. *Nat. Methods* **2021**, *18*, 389−396.
(182) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(183) Li, Y.; Fang, J. PROTS-RF: A Robust Model for Predicting Mutation-Induced Protein Stability Changes. *PLoS One* **2012**, *7*, e47247.
(184) Jia, L.; Yarlagadda, R.; Reed, C. C. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. *PLoS One* **2015**, *10*, e0138022.
(185) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273−297.
(186) Capriotti, E.; Fariselli, P.; Calabrese, R.; Casadio, R. Predicting Protein Stability Changes from Sequences Using Support Vector Machines. *Bioinformatics* **2005**, *21*, 54−58.
(187) Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins: Struct. Funct. Genet.* **2006**, *62*, 1125−1132.
(188) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant 2.0: Predicting Stability Changes Upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306−W310.
(189) Saladi, S. M.; Javed, N.; Müller, A.; Clemons, W. M., Jr A Statistical Model for Improved Membrane Protein Expression Using Sequence-Derived Features. *J. Biol. Chem.* **2018**, *293*, 4913−4927.
(190) Rasmussen, C. E. In *Advanced Lectures on Machine Learning*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer: Berlin, 2004.
(191) Jokinen, E.; Heinonen, M.; Lähdesmäki, H. mGPfusion: Predicting Protein Stability Changes with Gaussian Process Kernel Learning and Data Fusion. *Bioinformatics* **2018**, *34*, i274−i283.
(192) Pires, D. E.; Ascher, D. B.; Blundell, T. L. mCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* **2014**, *30*, 335−342.
(193) Mellor, J.; Grigoras, I.; Carbonell, P.; Faulon, J.-L. Semi-supervised Gaussian Process for Automated Enzyme Search. *ACS Synth. Biol.* **2016**, *5*, 518−528.
(194) Saito, Y.; Oikawa, M.; Nakazawa, H.; Niide, T.; Kameda, T.; Tsuda, K.; Umetsu, M. Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins. *ACS Synth. Biol.* **2018**, *7*, 2014−2022.
(195) Alipanahi, B.; Delong, A.; Weirauch, M. T.; Frey, B. J. Predicting the Sequence Specificities of DNA- and RNA-Binding Proteins by Deep Learning. *Nat. Biotechnol.* **2015**, *33*, 831−838.
(196) Zeng, H.; Edwards, M. D.; Liu, G.; Gifford, D. K. Convolutional Neural Network Architectures for Predicting DNA-Protein Binding. *Bioinformatics* **2016**, *32*, i121−i127.
(197) Zhang, S.; Zhou, J.; Hu, H.; Gong, H.; Chen, L.; Cheng, C.; Zeng, J. A Deep Learning Framework for Modeling Structural

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Features of RNA-Binding Protein Targets. *Nucleic Acids Res.* **2016**, *44*, No. e32.
(198) Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A. S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33*, 3036−3042.
(199) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and Accurate Predictions of Protein Stability Changes Upon Mutations Using Statistical Potentials and Neural Networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25*, 2537−2543.
(200) Khurana, S.; Rawi, R.; Kunji, K.; Chuang, G. Y.; Bensmail, H.; Mall, R. DeepSol: A Deep Learning Framework for Sequence-Based Protein Solubility Prediction. *Bioinformatics* **2018**, *34*, 2605−2613.
(201) Hopf, T. A.; Colwell, L. J.; Sheridan, R.; Rost, B.; Sander, C.; Marks, D. S. Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell* **2012**, *149*, 1607−1621.
(202) Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434* **2015**.
(203) Hopf, T. A.; Ingraham, J. B.; Poelwijk, F. J.; Schärfe, C. P. I.; Springer, M.; Sander, C.; Marks, D. S. Mutation Effects Predicted from Sequence Co-Variation. *Nat. Biotechnol.* **2017**, *35*, 128−135.
(204) Riesselman, A. J.; Ingraham, J. B.; Marks, D. S. Deep Generative Models of Genetic Variation Capture the Effects of Mutations. *Nat. Methods* **2018**, *15*, 816−822.
(205) Snoek, J.; Larochelle, H.; Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. *NIPS Proc.* **2012**, 2951−2959.
(206) Brookes, D.; Park, H.; Listgarten, J. Conditioning by Adaptive Sampling for Robust Design. *Int. Conf. Machine Learning* **2019**, 773−782.
(207) Wolpert, D. H. The Lack of a Priori Distinctions between Learning Algorithms. *Neural. Comput.* **1996**, *8*, 1341−1390.
(208) Li, M. Applications of Display Technology in Protein Analysis. *Nat. Biotechnol.* **2000**, *18*, 1251−1256.
(209) Demartis, S.; Huber, A.; Viti, F.; Lozzi, L.; Giovannoni, L.; Neri, P.; Winter, G.; Neri, D. A Strategy for the Isolation of Catalytic Activities from Repertoires of Enzymes Displayed on Phage. *J. Mol. Biol.* **1999**, *286*, 617−633.
(210) Jestin, J. L.; Kristensen, P.; Winter, G. A Method for the Selection of Catalytic Activity Using Phage Display and Proximity Coupling. *Angew. Chem., Int. Ed.* **1999**, *38*, 1124−1127.
(211) Paschke, M. Phage Display Systems and Their Applications. *Appl. Microbiol. Biotechnol.* **2006**, *70*, 2−11.
(212) Pedersen, H.; Holder, S.; Sutherlin, D. P.; Schwitter, U.; King, D. S.; Schultz, P. G. A Method for Directed Evolution and Functional Cloning of Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 10523−10528.
(213) Smith, G. P. Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228*, 1315−1317.
(214) Urban, J. H.; Merten, C. A. Retroviral Display in Gene Therapy, Protein Engineering, and Vaccine Development. *ACS Chem. Biol.* **2011**, *6*, 61−74.
(215) Granieri, L.; Baret, J. C.; Griffiths, A. D.; Merten, C. A. High-Throughput Screening of Enzymes by Retroviral Display Using Droplet-Based Microfluidics. *Chem. Biol.* **2010**, *17*, 229−235.
(216) Lee, S. Y.; Choi, J. H.; Xu, Z. Microbial Cell-Surface Display. *Trends Biotechnol.* **2003**, *21*, 45−52.
(217) Becker, S.; Schmoldt, H. U.; Adams, T. M.; Wilhelm, S.; Kolmar, H. Ultra-High-Throughput Screening Based on Cell-Surface Display and Fluorescence-Activated Cell Sorting for the Identification of Novel Biocatalysts. *Curr. Opin. Biotechnol.* **2004**, *15*, 323−329.
(218) Kim, Y. S.; Jung, H. C.; Pan, J. G. Bacterial Cell Surface Display of an Enzyme Library for Selective Screening of Improved Cellulase Variants. *Appl. Environ. Microbiol.* **2000**, *66*, 788−793.
(219) Boder, E. T.; Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nat. Biotechnol.* **1997**, *15*, 553−557.
(220) Cherf, G. M.; Cochran, J. R. Applications of Yeast Surface Display for Protein Engineering. *Methods Mol. Biol.* **2015**, *1319*, 155−175.
(221) Chen, K. C.; Wu, C. H.; Chang, C. Y.; Lu, W. C.; Tseng, Q.; Prijovich, Z. M.; Schechinger, W.; Liaw, Y. C.; Leu, Y. L.; Roffler, S. R. Directed Evolution of a Lysosomal Enzyme with Enhanced Activity at Neutral pH by Mammalian Cell-Surface Display. *Chem. Biol.* **2008**, *15*, 1277−1286.
(222) Yi, L.; Gebhard, M. C.; Li, Q.; Taft, J. M.; Georgiou, G.; Iverson, B. L. Engineering of Tev Protease Variants by Yeast ER Sequestration Screening (YESS) of Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7229−7234.
(223) 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. U. S. A.* **2011**, *108*, 11399−11404.
(224) Choi, Y. S.; Pack, S. P.; Yoo, Y. J. Development of a Plasmid Display System Using GAL4 DNA Binding Domain for the in Vitro Screening of Functional Proteins. *Biotechnol. Lett.* **2005**, *27*, 1707−1711.
(225) Hanes, J.; Pluckthun, A. In Vitro Selection and Evolution of Functional Proteins by Using Ribosome Display. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4937−4942.
(226) Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. In Vitro Virus: Bonding of mRNA Bearing Puromycin at the 3′-Terminal End to the C-Terminal End of Its Encoded Protein on the Ribosome in Vitro. *FEBS Lett.* **1997**, *414*, 405−408.
(227) Roberts, R. W.; Szostak, J. W. RNA-Peptide Fusions for the in Vitro Selection of Peptides and Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12297−12302.
(228) Valencia, C. A.; Zou, J.; Liu, R. In Vitro Selection of Proteins with Desired Characteristics Using mRNA-Display. *Methods* **2013**, *60*, 55−69.
(229) Houlihan, G.; Gatti-Lafranconi, P.; Kaltenbach, M.; Lowe, D.; Hollfelder, F. An Experimental Framework for Improved Selection of Binding Proteins Using SNAP Display. *J. Immunol. Methods* **2014**, *405*, 47−56.
(230) Kaltenbach, M.; Stein, V.; Hollfelder, F. SNAP Dendrimers: Multivalent Protein Display on Dendrimer-Like DNA for Directed Evolution. *ChemBioChem* **2011**, *12*, 2208−2216.
(231) Miller, B. G.; Raines, R. T. Identifying Latent Enzyme Activities: Substrate Ambiguity within Modern Bacterial Sugar Kinases. *Biochemistry* **2004**, *43*, 6387−6392.
(232) Boersma, Y. L.; Droge, M. J.; van der Sloot, A. M.; Pijning, T.; Cool, R. H.; Dijkstra, B. W.; Quax, W. J. A Novel Genetic Selection System for Improved Enantioselectivity of Bacillus Subtilis Lipase A. *ChemBioChem* **2008**, *9*, 1110−1115.
(233) van Sint Fiet, S.; van Beilen, J. B.; Witholt, B. Selection of Biocatalysts for Chemical Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1693−1698.
(234) Baker, K.; Bleczinski, C.; Lin, H.; Salazar-Jimenez, G.; Sengupta, D.; Krane, S.; Cornish, V. W. Chemical Complementation: A Reaction-Independent Genetic Assay for Enzyme Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16537−16542.
(235) Lin, H.; Tao, H.; Cornish, V. W. Directed Evolution of a Glycosynthase Via Chemical Complementation. *J. Am. Chem. Soc.* **2004**, *126*, 15051−15059.
(236) van den Berg, S.; Löfdahl, P.-Å.; Härd, T.; Berglund, H. Improved Solubility of TEV Protease by Directed Evolution. *J. Biotechnol.* **2006**, *121*, 291−298.
(237) Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. Rapid Protein-Folding Assay Using Green Fluorescent Protein. *Nat. Biotechnol.* **1999**, *17*, 691−695.
(238) Tomatis, P. E.; Rasia, R. M.; Segovia, L.; Vila, A. J. Mimicking Natural Evolution in Metallo-B-Lactamases through Second-Shell Ligand Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13761−13766.
(239) McLoughlin, S. Y.; Jackson, C.; Liu, J. W.; Ollis, D. L. Growth of Escherichia coli Coexpressing Phosphotriesterase and Glycer-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

ophosphodiester Phosphodiesterase, Using Paraoxon as the Sole Phosphorus Source. *Appl. Environ. Microbiol.* **2004**, *70*, 404−412.
(240) Sakamoto, T.; Joern, J. M.; Arisawa, A.; Arnold, F. H. Laboratory Evolution of Toluene Dioxygenase to Accept 4-Picoline as a Substrate. *Appl. Environ. Microbiol.* **2001**, *67*, 3882−3887.
(241) Colas, P.; Brent, R. The Impact of Two-Hybrid and Related Methods on Biotechnology. *Trends Biotechnol.* **1998**, *16*, 355−363.
(242) Brachmann, R. K.; Boeke, J. D. Tag Games in Yeast: The Two-Hybrid System and Beyond. *Curr. Opin. Biotechnol.* **1997**, *8*, 561−568.
(243) Lin, H.; Cornish, V. W. In Vivo Protein-Protein Interaction Assays: Beyond Proteins. *Angew. Chem., Int. Ed.* **2001**, *40*, 871−875.
(244) McNabb, D. S.; Guarente, L. Genetic and Biochemical Probes for Protein-Protein Interactions. *Curr. Opin. Biotechnol.* **1996**, *7*, 554−559.
(245) Vidal, M.; Legrain, P. Yeast Forward and Reverse 'N'-Hybrid Systems. *Nucleic Acids Res.* **1999**, *27*, 919−929.
(246) Warbrick, E. Two's Company, Three's a Crowd: The Yeast Two Hybrid System for Mapping Molecular Interactions. *Structure* **1997**, *5*, 13−17.
(247) Drees, B. L. Progress and Variations in Two-Hybrid and Three-Hybrid Technologies. *Curr. Opin. Chem. Biol.* **1999**, *3*, 64−70.
(248) Tawfik, D. S.; Griffiths, A. D. Man-Made Cell-Like Compartments for Molecular Evolution. *Nat. Biotechnol.* **1998**, *16*, 652−656.
(249) Aharoni, A.; Gaidukov, L.; Yagur, S.; Toker, L.; Silman, I.; Tawfik, D. S. Directed Evolution of Mammalian Paraoxonases PON1 and PON3 for Bacterial Expression and Catalytic Specialization. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 482−487.
(250) Mastrobattista, E.; Taly, V.; Chanudet, E.; Treacy, P.; Kelly, B. T.; Griffiths, A. D. High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a Beta-Galactosidase by Fluorescence-Activated Sorting of Double Emulsions. *Chem. Biol.* **2005**, *12*, 1291−1300.
(251) Stapleton, J. A.; Swartz, J. R. Development of an in Vitro Compartmentalization Screen for High-Throughput Directed Evolution of [FeFe] Hydrogenases. *PLoS One* **2010**, *5*, No. e15275.
(252) Wu, C. H.; Liu, I. J.; Lu, R. M.; Wu, H. C. Advancement and Applications of Peptide Phage Display Technology in Biomedical Science. *J. Biomed. Sci.* **2016**, *23*, 8.
(253) Khurana, S.; Suguitan, A. L.; Rivera, Y.; Simmons, C. P.; Lanzavecchia, A.; Sallusto, F.; Manischewitz, J.; King, L. R.; Subbarao, K.; Golding, H. Antigenic Fingerprinting of H5N1 Avian Influenza Using Convalescent Sera and Monoclonal Antibodies Reveals Potential Vaccine and Diagnostic Targets. *PLoS Med.* **2009**, *6*, 6.
(254) Liu, I. J.; Hsueh, P. R.; Lin, C. T.; Chiu, C. Y.; Kao, C. L.; Liao, M. Y.; Wu, H. C. Disease-Specific B Cell Epitopes for Serum Antibodies from Patients with Severe Acute Respiratory Syndrome (SARS) and Serologic Detection of SARS Antibodies by Epitope-Based Peptide Antigens. *J. Infect. Dis.* **2004**, *190*, 797−809.
(255) Santamaria, H.; Manoutcharian, K.; Rocha, L.; Gonzalez, E.; Acero, G.; Govezensky, T.; Uribe, L. I.; Olguin, A.; Paniagua, J.; Gevorkian, G. Identification of Peptide Sequences Specific for Serum Antibodies from Human Papillomavirus-Infected Patients Using Phage Display Libraries. *Clin. Immunol.* **2001**, *101*, 296−302.
(256) Lunardi, C.; Bason, C.; Navone, R.; Millo, E.; Damonte, G.; Corrocher, R.; Puccetti, A. Systemic Sclerosis Immunoglobulin G Autoantibodies Bind the Human Cytomegalovirus Late Protein UL94 and Induce Apoptosis in Human Endothelial Cells. *Nat. Med.* **2000**, *6*, 1183−1186.
(257) Yu, R. K.; Usuki, S.; Itokazu, Y.; Wu, H. C. Novel GM1 Ganglioside-Like Peptide Mimics Prevent the Association of Cholera Toxin to Human Intestinal Epithelial in Vitro. *Glycobiology* **2016**, *26*, 540.
(258) Arap, W.; Kolonin, M. G.; Trepel, M.; Lahdenranta, J.; Cardo-Vila, M.; Giordano, R. J.; Mintz, P. J.; Ardelt, P. U.; Yao, V. J.; Vidal, C. I.; et al. Steps toward Mapping the Human Vasculature by Phage Display. *Nat. Med.* **2002**, *8*, 121−127.
(259) Chang, D. K.; Lin, C. T.; Wu, C. H.; Wu, H. C. A Novel Peptide Enhances Therapeutic Efficacy of Liposomal Anti-Cancer Drugs in Mice Models of Human Lung Cancer. *PLoS One* **2009**, *4*, No. e4171.
(260) Krag, D. N.; Shukla, G. S.; Shen, G. P.; Pero, S.; Ashikaga, T.; Fuller, S.; Weaver, D. L.; Burdette-Radoux, S.; Thomas, C. Selection of Tumor-Binding Ligands in Cancer Patients with Phage Display Libraries. *Cancer Res.* **2006**, *66*, 7724−7733.
(261) Lee, T. Y.; Lin, C. T.; Kuo, S. Y.; Chang, D. K.; Wu, H. C. Peptide-Mediated Targeting to Tumor Blood Vessels of Lung Cancer for Drug Delivery. *Cancer Res.* **2007**, *67*, 10958−10965.
(262) Rajotte, D.; Arap, W.; Hagedorn, M.; Koivunen, E.; Pasqualini, R.; Ruoslahti, E. Molecular Heterogeneity of the Vascular Endothelium Revealed by in Vivo Phage Display. *J. Clin. Invest.* **1998**, *102*, 430−437.
(263) Gray, B. P.; Brown, K. C. Combinatorial Peptide Libraries: Mining for Cell-Binding Peptides. *Chem. Rev.* **2014**, *114*, 1020−1081.
(264) Lee, T. Y.; Wu, H. C.; Tseng, Y. L.; Lin, C. T. A Novel Peptide Specifically Binding to Nasopharyngeal Carcinoma for Targeted Drug Delivery. *Cancer Res.* **2004**, *64*, 8002−8008.
(265) Lu, R. M.; Chen, M. S.; Chang, D. K.; Chiu, C. Y.; Lin, W. C.; Yan, S. L.; Wang, Y. P.; Kuo, Y. S.; Yeh, C. Y.; Lo, A.; et al. Targeted Drug Delivery Systems Mediated by a Novel Peptide in Breast Cancer Therapy and Imaging. *PLoS One* **2013**, *8*, No. e66128.
(266) Wu, C. H.; Kuo, Y. H.; Hong, R. L.; Wu, H. C. Alpha-Enolase-Binding Peptide Enhances Drug Delivery Efficiency and Therapeutic Efficacy against Colorectal Cancer. *Sci. Transl. Med.* **2015**, *7*, 290ra91.
(267) Giordano, R. J.; Cardo-Vila, M.; Lahdenranta, J.; Pasqualini, R.; Arap, W. Biopanning and Rapid Analysis of Selective Interactive Ligands. *Nat. Med.* **2001**, *7*, 1249−1253.
(268) Lo, A.; Lin, C. T.; Wu, H. C. Hepatocellular Carcinoma Cell-Specific Peptide Ligand for Targeted Drug Delivery. *Mol. Cancer Ther.* **2008**, *7*, 579−589.
(269) McGuire, M. J.; Li, S.; Brown, K. C. Biopanning of Phage Displayed Peptide Libraries for the Isolation of Cell-Specific Ligands. *Methods Mol. Biol.* **2009**, *504*, 291−321.
(270) Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. Selection of Peptides with Semiconductor Binding Specificity for Directed Nanocrystal Assembly. *Nature* **2000**, *405*, 665−668.
(271) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818−822.
(272) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249*, 505−510.
(273) Stoltenburg, R.; Reinemann, C.; Strehlitz, B. SELEX–a (R)Evolutionary Method to Generate High-Affinity Nucleic Acid Ligands. *Biomol. Eng.* **2007**, *24*, 381−403.
(274) Adachi, T.; Nakamura, Y. Aptamers: A Review of Their Chemical Properties and Modifications for Therapeutic Application. *Molecules* **2019**, *24*, 4229.
(275) Gragoudas, E. S.; Adamis, A. P.; Cunningham, E. T., Jr; Feinsod, M.; Guyer, D. R. Pegaptanib for Neovascular Age-Related Macular Degeneration. *N. Engl. J. Med.* **2004**, *351*, 2805−2816.
(276) Takeda, A. L.; Colquitt, J.; Clegg, A. J.; Jones, J. Pegaptanib and Ranibizumab for Neovascular Age-Related Macular Degeneration: A Systematic Review. *Br. J. Ophthalmol.* **2007**, *91*, 1177−1182.
(277) Trujillo, C. A.; Nery, A. A.; Alves, J. M.; Martins, A. H.; Ulrich, H. Development of the Anti-VEGF Aptamer to a Therapeutic Agent for Clinical Ophthalmology. *Clin. Ophthalmol.* **2007**, *1*, 393−402.
(278) Cossarizza, A.; Chang, H. D.; Radbruch, A.; Acs, A.; Adam, D.; Adam-Klages, S.; Agace, W. W.; Aghaeepour, N.; Akdis, M.; Allez, M.; et al. Guidelines for the Use of Flow Cytometry and Cell Sorting in Immunological Studies (Second Edition). *Eur. J. Immunol.* **2019**, *49*, 1457−1973.
(279) Ibrahim, S. F.; van den Engh, G. Flow Cytometry and Cell Sorting. *Adv. Biochem. Eng. Biotechnol.* **2007**, *106*, 19−39.
(280) Yang, G.; Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **2009**, *10*, 2704−2715.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(281) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54*, 4011−4020.
(282) van Rossum, T.; Kengen, S. W.; van der Oost, J. Reporter-Based Screening and Selection of Enzymes. *FEBS J.* **2013**, *280*, 2979−2996.
(283) Michener, J. K.; Smolke, C. D. High-Throughput Enzyme Evolution in *Saccharomyces cerevisiae* Using a Synthetic RNA Switch. *Metab. Eng.* **2012**, *14*, 306−316.
(284) Huber, R.; Wulfhorst, H.; Maksym, L.; Stehr, R.; Pohnlein, M.; Jager, G.; Spiess, A. C.; Buchs, J. Screening for Enzyme Activity in Turbid Suspensions with Scattered Light. *Biotechnol. Prog.* **2011**, *27*, 555−561.
(285) Jager, G.; Wulfhorst, H.; Zeithammel, E. U.; Elinidou, E.; Spiess, A. C.; Buchs, J. Screening of Cellulases for Biofuel Production: Online Monitoring of the Enzymatic Hydrolysis of Insoluble Cellulose Using High-Throughput Scattered Light Detection. *Biotechnol. J.* **2011**, *6*, 74−85.
(286) Mack, M.; Burger, M.; Pietschmann, P.; Hock, B. A High-Throughput Microtiter Plate-Based Screening Method for the Detection of Full-Length Recombinant Proteins. *Protein Expression Purif.* **2008**, *61*, 92−98.
(287) Duetz, W. A. Microtiter Plates as Mini-Bioreactors: Miniaturization of Fermentation Methods. *Trends Microbiol.* **2007**, *15*, 469−475.
(288) Topolska, M.; Martinez-Montanes, F.; Ejsing, C. S. A Simple and Direct Assay for Monitoring Fatty Acid Synthase Activity and Product-Specificity by High-Resolution Mass Spectrometry. *Biomolecules* **2020**, *10*, 118.
(289) de Rond, T.; Danielewicz, M.; Northen, T. High Throughput Screening of Enzyme Activity with Mass Spectrometry Imaging. *Curr. Opin. Biotechnol.* **2015**, *31*, 1−9.
(290) Marcellin, E.; Nielsen, L. K. Advances in Analytical Tools for High Throughput Strain Engineering. *Curr. Opin. Biotechnol.* **2018**, *54*, 33−40.
(291) Silva, L. P.; Northen, T. R. Exometabolomics and MSI: Deconstructing How Cells Interact to Transform Their Small Molecule Environment. *Curr. Opin. Biotechnol.* **2015**, *34*, 209−216.
(292) Xue, P.; Si, T.; Mishra, S.; Zhang, L.; Choe, K.; Sweedler, J. V.; Zhao, H. A Mass Spectrometry-Based High-Throughput Screening Method for Engineering Fatty Acid Synthases with Improved Production of Medium-Chain Fatty Acids. *Biotechnol. Bioeng.* **2020**, *117*, 2131−2138.
(293) Gray, C. J.; Weissenborn, M. J.; Eyers, C. E.; Flitsch, S. L. Enzymatic Reactions on Immobilised Substrates. *Chem. Soc. Rev.* **2013**, *42*, 6378−6405.
(294) Si, T.; Li, B.; Comi, T. J.; Wu, Y.; Hu, P.; Wu, Y.; Min, Y.; Mitchell, D. A.; Zhao, H.; Sweedler, J. V. Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions Via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **2017**, *139*, 12466−12473.
(295) Rubakhin, S. S.; Garden, R. W.; Fuller, R. R.; Sweedler, J. V. Measuring the Peptides in Individual Organelles with Mass Spectrometry. *Nat. Biotechnol.* **2000**, *18*, 172−175.
(296) Nemes, P.; Rubakhin, S. S.; Aerts, J. T.; Sweedler, J. V. Qualitative and Quantitative Metabolomic Investigation of Single Neurons by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry. *Nat. Protoc.* **2013**, *8*, 783−799.
(297) Xue, P.; Si, T.; Zhao, H. In *Methods Enzymol.*; Tawfik, D. S., Ed.; Academic Press, 2020; Vol. *644*.
(298) Mair, P.; Gielen, F.; Hollfelder, F. Exploring Sequence Space in Search of Functional Enzymes Using Microfluidic Droplets. *Curr. Opin. Chem. Biol.* **2017**, *37*, 137−144.
(299) Liu, W.-w.; Zhu, Y. Development and Application of Analytical Detection Techniques for Droplet-Based Microfluidics"-a Review. *Anal. Chim. Acta* **2020**, *1113*, 66−84.
(300) Fidalgo, L. M.; Whyte, G.; Ruotolo, B. T.; Benesch, J. L. P.; Stengel, F.; Abell, C.; Robinson, C. V.; Huck, W. T. S. Coupling Microdroplet Microreactors with Mass Spectrometry: Reading the Contents of Single Droplets Online. *Angew. Chem., Int. Ed.* **2009**, *48*, 3665−3668.
(301) Pei, J.; Li, Q.; Lee, M. S.; Valaskovic, G. A.; Kennedy, R. T. Analysis of Samples Stored as Individual Plugs in a Capillary by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 6558−6561.
(302) Hatakeyama, T.; Chen, D. L.; Ismagilov, R. F. Microgram-Scale Testing of Reaction Conditions in Solution Using Nanoliter Plugs in Microfluidics with Detection by MALDI-MS. *J. Am. Chem. Soc.* **2006**, *128*, 2518−2519.
(303) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J.-C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. A. Ultrahigh-Throughput Screening in Drop-Based Microfluidics for Directed Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4004−4009.
(304) Mao, Z.; Guo, F.; Xie, Y.; Zhao, Y.; Lapsley, M. I.; Wang, L.; Mai, J. D.; Costanzo, F.; Huang, T. J. Label-Free Measurements of Reaction Kinetics Using a Droplet-Based Optofluidic Device. *J. Lab. Autom.* **2015**, *20*, 17−24.
(305) Braun, F.; Schwolow, S.; Seltenreich, J.; Kockmann, N.; Röder, T.; Gretz, N.; Rädle, M. Highly Sensitive Raman Spectroscopy with Low Laser Power for Fast in-Line Reaction and Multiphase Flow Monitoring. *Anal. Chem.* **2016**, *88*, 9368−9374.
(306) Suea-Ngam, A.; Rattanarat, P.; Chailapakul, O.; Srisa-Art, M. Electrochemical Droplet-Based Microfluidics Using Chip-Based Carbon Paste Electrodes for High-Throughput Analysis in Pharmaceutical Applications. *Anal. Chim. Acta* **2015**, *883*, 45−54.
(307) Kautz, R. A.; Goetzinger, W. K.; Karger, B. L. High-Throughput Microcoil NMR of Compound Libraries Using Zero-Dispersion Segmented Flow Analysis. *J. Comb. Chem.* **2005**, *7*, 14−20.
(308) Diefenbach, X. W.; Farasat, I.; Guetschow, E. D.; Welch, C. J.; Kennedy, R. T.; Sun, S. W.; Moore, J. C. Enabling Biocatalysis by High-Throughput Protein Engineering Using Droplet Microfluidics Coupled to Mass Spectrometry. *ACS Omega* **2018**, *3*, 1498−1508.
(309) Colin, P. Y.; Kintses, B.; Gielen, F.; Miton, C. M.; Fischer, G.; Mohamed, M. F.; Hyvonen, M.; Morgavi, D. P.; Janssen, D. B.; Hollfelder, F. Ultrahigh-Throughput Discovery of Promiscuous Enzymes by Picodroplet Functional Metagenomics. *Nat. Commun.* **2015**, *6*, 10008.
(310) Bylina, E. J.; Grek, C. L.; Coleman, W. J.; Youvan, D. C. Directed Evolution and Solid Phase Enzyme Screening. *Advances in Nucleic Acid and Protein Analyses, Manipulation, and Sequencing* **2000**, *1*, 186−191.
(311) Delagrave, S.; Murphy, D. J.; Pruss, J. L. R.; Maffia, A. M., III; Marrs, B. L.; Bylina, E. J.; Coleman, W. J.; Grek, C. L.; Dilworth, M. R.; Yang, M. M.; et al. Application of a Very High-Throughput Digital Imaging Screen to Evolve the Enzyme Galactose Oxidase. *Protein Eng., Des. Sel.* **2001**, *14*, 261−267.
(312) Kone, F. M.; Le Bechec, M.; Sine, J. P.; Dion, M.; Tellier, C. Digital Screening Methodology for the Directed Evolution of Transglycosidases. *Protein Eng., Des. Sel.* **2008**, *22*, 37−44.
(313) Avoigt, C.; Kauffman, S.; Wang, Z.-G. In *Advances in Protein Chemistry*; Academic Press, 2001; Vol. *55*.
(314) Wright, M. C.; Joyce, G. F. Continuous in Vitro Evolution of Catalytic Function. *Science* **1997**, *276*, 614−617.
(315) Joyce, G. F. Directed Evolution of Nucleic Acid Enzymes. *Annu. Rev. Biochem.* **2004**, *73*, 791−836.
(316) Luan, G.; Cai, Z.; Li, Y.; Ma, Y. Genome Replication Engineering Assisted Continuous Evolution (GREACE) to Improve Microbial Tolerance for Biofuels Production. *Biotechnol. Biofuels* **2013**, *6*, 137.
(317) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472*, 499−503.
(318) Fijalkowska, I. J.; Schaaper, R. M. Mutants in the Exo I Motif of *Escherichia coli* dnaQ: Defective Proofreading and Inviability Due to Error Catastrophe. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 2856−2861.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(319) Opperman, T.; Murli, S.; Smith, B. T.; Walker, G. C. A Model for a *UmuDC*-Dependent Prokaryotic DNA Damage Checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9218−9223.
(320) 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.* **2014**, *10*, 216−222.
(321) Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a Split RNA Polymerase as a Versatile Biosensor Platform. *Nat. Chem. Biol.* **2017**, *13*, 432−438.
(322) Dickinson, B. C.; Leconte, A. M.; Allen, B.; Esvelt, K. M.; Liu, D. R. Experimental Interrogation of the Path Dependence and Stochasticity of Protein Evolution Using Phage-Assisted Continuous Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 9007−9012.
(323) Dove, S. L.; Hochschild, A. Conversion of the Omega Subunit of *Escherichia coli* RNA Polymerase into a Transcriptional Activator or an Activation Target. *Genes Dev.* **1998**, *12*, 745−754.
(324) Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M.; Chen, L.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R. Continuous Directed Evolution of DNA-Binding Proteins to Improve TALEN Specificity. *Nat. Methods* **2015**, *12*, 939−942.
(325) Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; et al. Evolved Cas9 Variants with Broad PAM Compatibility and High DNA Specificity. *Nature* **2018**, *556*, 57−63.
(326) Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A. M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; et al. Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature* **2016**, *533*, 58−63.
(327) Wang, T.; Badran, A. H.; Huang, T. P.; Liu, D. R. Continuous Directed Evolution of Proteins with Improved Soluble Expression. *Nat. Chem. Biol.* **2018**, *14*, 972−980.
(328) 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.* **2014**, *5*, 5352.
(329) Packer, M. S.; Rees, H. A.; Liu, D. R. Phage-Assisted Continuous Evolution of Proteases with Altered Substrate Specificity. *Nat. Commun.* **2017**, *8*, 956.
(330) Bryson, D. I.; Fan, C.; Guo, L.-T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-tRNA Synthetases. *Nat. Chem. Biol.* **2017**, *13*, 1253−1260.
(331) Berman, C. M.; Papa, L. J., 3rd; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; et al. An Adaptable Platform for Directed Evolution in Human Cells. *J. Am. Chem. Soc.* **2018**, *140*, 18093−18103.
(332) Bull, A. T. The Renaissance of Continuous Culture in the Post-Genomics Age. *J. Ind. Microbiol. Biotechnol.* **2010**, *37*, 993−1021.
(333) Gresham, D.; Dunham, M. J. The Enduring Utility of Continuous Culturing in Experimental Evolution. *Genomics* **2014**, *104*, 399−405.
(334) Piper, M. D.; Daran-Lapujade, P.; Bro, C.; Regenberg, B.; Knudsen, S.; Nielsen, J.; Pronk, J. T. Reproducibility of Oligonucleotide Microarray Transcriptome Analyses. An Interlaboratory Comparison Using Chemostat Cultures of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2002**, *277*, 37001−37008.
(335) Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* **2008**, *383*, 1069−1080.
(336) Wong, B. G.; Mancuso, C. P.; Kiriakov, S.; Bashor, C. J.; Khalil, A. S. Precise, Automated Control of Conditions for High-Throughput Growth of Yeast and Bacteria with eVOLVER. *Nat. Biotechnol.* **2018**, *36*, 614−623.
(337) Zhong, Z.; Wong, B. G.; Ravikumar, A.; Arzumanyan, G. A.; Khalil, A. S.; Liu, C. C. Automated Continuous Evolution of Proteins in Vivo. *ACS Synth. Biol.* **2020**, *9*, 1270−1276.
(338) Ellington, A. D.; Szostak, J. W. Invitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818−822.
(339) Landweber, L. F. Experimental RNA Evolution. *Trends Ecol. Evol.* **1999**, *14*, 353−358.
(340) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16*, 379−394.
(341) Schmidt-Dannert, C. Directed Evolution of Single Proteins, Metabolic Pathways, and Viruses. *Biochemistry* **2001**, *40*, 13125−13136.
(342) Schimmel, P.; Kelley, S. O. Exiting an RNA World. *Nat. Struct. Biol.* **2000**, *7*, 5−7.
(343) Ordoukhanian, P.; Joyce, G. F. A Molecular Description of the Evolution of Resistance. *Chem. Biol.* **1999**, *6*, 881−889.
(344) Kuhne, H.; Joyce, G. F. Continuous in Vitro Evolution of Ribozymes That Operate under Conditions of Extreme pH. *J. Mol. Evol.* **2003**, *57*, 292−298.
(345) Miyamoto, Y.; Teramoto, N.; Imanishi, Y.; Ito, Y. In Vitro Adaptation of a Ligase Ribozyme for Activity under a Low-pH Condition. *Biotechnol. Bioeng.* **2001**, *75*, 590−596.
(346) Lehman, N. Assessing the Likelihood of Recurrence During RNA Evolution in Vitro. *Artif. Life* **2004**, *10*, 1−22.
(347) Santoro, S. W.; Joyce, G. F. A General Purpose RNA-Cleaving DNA Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4262−4266.
(348) Breaker, R. R.; Joyce, G. F. A DNA Enzyme That Cleaves RNA. *Chem. Biol.* **1994**, *1*, 223−229.
(349) Breaker, R. R.; Joyce, G. F. A DNA Enzyme with Mg(2+)-Dependent RNA Phosphoesterase Activity. *Chem. Biol.* **1995**, *2*, 655−660.
(350) Li, J.; Lu, Y. A Highly Sensitive and Selective Catalytic DNA Biosensor for Lead Ions. *J. Am. Chem. Soc.* **2000**, *122*, 10466−10467.
(351) Liu, J.; Lu, Y. A Colorimetric Lead Biosensor Using DNAzyme-Directed Assembly of Gold Nanoparticles. *J. Am. Chem. Soc.* **2003**, *125*, 6642−6643.
(352) Zhang, L.; Gasper, W. J.; Stass, S. A.; Ioffe, O. B.; Davis, M. A.; Mixson, A. J. Angiogenic Inhibition Mediated by a DNAzyme That Targets Vascular Endothelial Growth Factor Receptor 2. *Cancer Res.* **2002**, *62*, 5463−5469.
(353) Weigand, J. E.; Suess, B. Aptamers and Riboswitches: Perspectives in Biotechnology. *Appl. Microbiol. Biotechnol.* **2009**, *85*, 229−236.
(354) Bastet, L.; Dubé, A.; Massé, E.; Lafontaine, D. A. New Insights into Riboswitch Regulation Mechanisms. *Mol. Microbiol.* **2011**, *80*, 1148−1154.
(355) Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W. Conditional Gene Expression by Controlling Translation with Tetracycline-Binding Aptamers. *Nucleic Acids Res.* **2003**, *31*, 1853−1858.
(356) Page, K.; Shaffer, J.; Lin, S.; Zhang, M.; Liu, J. M. Engineering Riboswitches in Vivo Using Dual Genetic Selection and Fluorescence-Activated Cell Sorting. *ACS Synth. Biol.* **2018**, *7*, 2000−2006.
(357) Pang, Q.; Han, H.; Liu, X.; Wang, Z.; Liang, Q.; Hou, J.; Qi, Q.; Wang, Q. In Vivo Evolutionary Engineering of Riboswitch with High-Threshold for N-Acetylneuraminic Acid Production. *Metab. Eng.* **2020**, *59*, 36−43.
(358) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in *Bacillus Subtilis* and Other Bacteria. *Cell* **2003**, *113*, 577−586.
(359) Mehrshahi, P.; Nguyen, G. T. D. T.; Gorchs Rovira, A.; Sayer, A.; Llavero-Pasquina, M.; Lim Huei Sin, M.; Medcalf, E. J.; Mendoza-Ochoa, G. I.; Scaife, M. A.; Smith, A. G. Development of Novel Riboswitches for Synthetic Biology in the Green Alga Chlamydomonas. *ACS Synth. Biol.* **2020**, *9*, 1406−1417.
(360) Boussebayle, A.; Torka, D.; Ollivaud, S.; Braun, J.; Bofill-Bosch, C.; Dombrowski, M.; Groher, F.; Hamacher, K.; Suess, B. Next-Level Riboswitch Development-Implementation of Capture-SELEX Facilitates Identification of a New Synthetic Riboswitch. *Nucleic Acids Res.* **2019**, *47*, 4883−4895.
(361) Eydeler, K.; Magbanua, E.; Werner, A.; Ziegelmüller, P.; Hahn, U. Fluorophore Binding Aptamers as a Tool for RNA Visualization. *Biophys. J.* **2009**, *96*, 3703−3707.
(362) Bai, J.; Luo, Y.; Wang, X.; Li, S.; Luo, M.; Yin, M.; Zuo, Y.; Li, G.; Yao, J.; Yang, H.; et al. A Protein-Independent Fluorescent RNA

Aptamer Reporter System for Plant Genetic Engineering. *Nat. Commun.* **2020**, *11*, 3847.
(363) Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J. Am. Chem. Soc.* **2014**, *136*, 16299−16308.
(364) Nechooshtan, G.; Elgrably-Weiss, M.; Sheaffer, A.; Westhof, E.; Altuvia, S. A pH-Responsive Riboregulator. *Genes Dev.* **2009**, *23*, 2650−2662.
(365) Pham, H. L.; Wong, A.; Chua, N.; Teo, W. S.; Yew, W. S.; Chang, M. W. Engineering a Riboswitch-Based Genetic Platform for the Self-Directed Evolution of Acid-Tolerant Phenotypes. *Nat. Commun.* **2017**, *8*, 411.
(366) Bornscheuer, U. T.; Hauer, B.; Jaeger, K. E.; Schwaneberg, U. Directed Evolution Empowered Redesign of Natural Proteins for the Sustainable Production of Chemicals and Pharmaceuticals. *Angew. Chem., Int. Ed.* **2019**, *58*, 36−40.
(367) Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87*, 131−157.
(368) Jäckel, C.; Kast, P.; Hilvert, D. Protein Design by Directed Evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153−173.
(369) Fasan, R.; Crook, N. C.; Peters, M. W.; Meinhold, P.; Buelter, T.; Landwehr, M.; Cirino, P. C.; Arnold, F. H. Improved Product-Per-Glucose Yields in P450-Dependent Propane Biotransformations Using Engineered *Escherichia coli*. *Biotechnol. Bioeng.* **2011**, *108*, 500−510.
(370) Fischbach, M. A.; Lai, J. R.; Roche, E. D.; Walsh, C. T.; Liu, D. R. Directed Evolution Can Rapidly Improve the Activity of Chimeric Assembly-Line Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11951−11956.
(371) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem., Int. Ed.* **2016**, *55*, 16110−16114.
(372) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H. Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6*, 7810−7813.
(373) Chen, K. Q.; Arnold, F. H. Enzyme Engineering for Nonaqueous Solvents: Random Mutagenesis to Enhance Activity of *Subtilisin E* in Polar Organic Media. *Bio/Technology* **1991**, *9*, 1073−1077.
(374) Moore, J. C.; Arnold, F. H. Directed Evolution of a Para-Nitrobenzyl Esterase for Aqueous-Organic Solvents. *Nat. Biotechnol.* **1996**, *14*, 458−467.
(375) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; et al. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305−309.
(376) Schallmey, A.; Schallmey, M. Recent Advances on Halohydrin Dehalogenases-from Enzyme Identification to Novel Biocatalytic Applications. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 7827−7839.
(377) de Jong, R. M.; Tiesinga, J. J. W.; Villa, A.; Tang, L.; Janssen, D. B.; Dijkstra, B. W. Structural Basis for the Enantioselectivity of an Epoxide Ring Opening Reaction Catalyzed by Halo Alcohol Dehalogenase HheC. *J. Am. Chem. Soc.* **2005**, *127*, 13338−13343.
(378) Davies, H. M. L.; Manning, J. R. Catalytic C−H Functionalization by Metal Carbenoid and Nitrenoid Insertion. *Nature* **2008**, *451*, 417−424.
(379) Lebel, H.; Marcoux, J.-F.; Molinaro, C.; Charette, A. B. Stereoselective Cyclopropanation Reactions. *Chem. Rev.* **2003**, *103*, 977−1050.
(380) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation Via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307−310.
(381) Lewis, R. D.; Garcia-Borràs, M.; Chalkley, M. J.; Buller, A. R.; Houk, K. N.; Kan, S. B. J.; Arnold, F. H. Catalytic Iron-Carbene Intermediate Revealed in a Cytochrome C Carbene Transferase. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 7308−7313.
(382) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A Serine-Substituted P450 Catalyzes Highly Efficient Carbene Transfer to Olefins in Vivo. *Nat. Chem. Biol.* **2013**, *9*, 485−487.
(383) Wang, Z. J.; Peck, N. E.; Renata, H.; Arnold, F. H. Cytochrome P450-Catalyzed Insertion of Carbenoids into N−H Bonds. *Chem. Sci.* **2014**, *5*, 598−601.
(384) Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102−111.
(385) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome C for Carbon−Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**, *354*, 1048−1051.
(386) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552*, 132−136.
(387) Huang, X.; Garcia-Borràs, M.; Miao, K.; Kan, S. B. J.; Zutshi, A.; Houk, K. N.; Arnold, F. H. A Biocatalytic Platform for Synthesis of Chiral A-Trifluoromethylated Organoborons. *ACS Cent. Sci.* **2019**, *5*, 270−276.
(388) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem., Int. Ed.* **2015**, *54*, 1744−1748.
(389) Sreenilayam, G.; Fasan, R. Myoglobin-Catalyzed Intermolecular Carbene N−H Insertion with Arylamine Substrates. *Chem. Commun.* **2015**, *51*, 1532−1534.
(390) Vargas, D. A.; Khade, R. L.; Zhang, Y.; Fasan, R. Biocatalytic Strategy for Highly Diastereo- and Enantioselective Synthesis of 2,3-Dihydrobenzofuran-Based Tricyclic Scaffolds. *Angew. Chem., Int. Ed.* **2019**, *58*, 10148−10152.
(391) Ren, X.; Liu, N.; Chandgude, A. L.; Fasan, R. An Enzymatic Platform for the Highly Enantioselective and Stereodivergent Construction of Cyclopropyl-*δ*-Lactones. *Angew. Chem., Int. Ed.* **2020**, *59*, 21634−21639.
(392) Liao, K.; Pickel, T. C.; Boyarskikh, V.; Bacsa, J.; Musaev, D. G.; Davies, H. M. L. Site-Selective and Stereoselective Functionalization of Non-Activated Tertiary C-H Bonds. *Nature* **2017**, *551*, 609−613.
(393) Weldy, N. M.; Schafer, A. G.; Owens, C. P.; Herting, C. J.; Varela-Alvarez, A.; Chen, S.; Niemeyer, Z.; Musaev, D. G.; Sigman, M. S.; Davies, H. M. L.; et al. Iridium(III)-Bis(Imidazolinyl)Phenyl Catalysts for Enantioselective C−H Functionalization with Ethyl Diazoacetate. *Chem. Sci.* **2016**, *7*, 3142−3146.
(394) Caballero, A.; Despagnet-Ayoub, E.; Mar Díaz-Requejo, M.; Díaz-Rodríguez, A.; González-Núñez, M. E.; Mello, R.; Muñoz, B. K.; Ojo, W.-S.; Asensio, G.; Etienne, M.; et al. Silver-Catalyzed C-C Bond Formation between Methane and Ethyl Diazoacetate in Supercritical CO_2. *Science* **2011**, *332*, 835−838.
(395) Mbuvi, H. M.; Woo, L. K. Catalytic C-H Insertions Using Iron(III) Porphyrin Complexes. *Organometallics* **2008**, *27*, 637−645.
(396) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon−Carbon Bonds Via Iron-Catalysed sp^3C−H Functionalization. *Nature* **2019**, *565*, 67−72.
(397) Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H. Enantiodivergent *α*-Amino C−H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, 9798−9802.
(398) Vargas, D. A.; Tinoco, A.; Tyagi, V.; Fasan, R. Myoglobin-Catalyzed C-H Functionalization of Unprotected Indoles. *Angew. Chem., Int. Ed.* **2018**, *57*, 9911−9915.
(399) Brandenberg, O. F.; Chen, K.; Arnold, F. H. Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J. Am. Chem. Soc.* **2019**, *141*, 8989−8995.
(400) Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. Enzyme-Controlled Nitrogen-Atom Transfer Enables

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Regiodivergent C−H Amination. *J. Am. Chem. Soc.* **2014**, *136*, 15505−15508.
(401) Singh, R.; Bordeaux, M.; Fasan, R. P450-Catalyzed Intramoleculars sp^3 C−H Amination with Arylsulfonyl Azide Substrates. *ACS Catal.* **2014**, *4*, 546−552.
(402) Singh, R.; Kolev, J. N.; Sutera, P. A.; Fasan, R. Enzymatic C(sp^3)-H Amination: P450-Catalyzed Conversion of Carbonazidates into Oxazolidinones. *ACS Catal.* **2015**, *5*, 1685−1691.
(403) Farwell, C. C.; McIntosh, J. A.; Hyster, T. K.; Wang, Z. J.; Arnold, F. H. Enantioselective Imidation of Sulfides Via Enzyme-Catalyzed Intermolecular Nitrogen-Atom Transfer. *J. Am. Chem. Soc.* **2014**, *136*, 8766−8771.
(404) Prier, C. K.; Hyster, T. K.; Farwell, C. C.; Huang, A.; Arnold, F. H. Asymmetric Enzymatic Synthesis of Allylic Amines: A Sigmatropic Rearrangement Strategy. *Angew. Chem., Int. Ed.* **2016**, *55*, 4711−4715.
(405) Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1*, 89−93.
(406) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C−H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nat. Chem.* **2017**, *9*, 629−634.
(407) Jia, Z.-J.; Gao, S.; Arnold, F. H. Enzymatic Primary Amination of Benzylic and Allylic C(sp^3)−H Bonds. *J. Am. Chem. Soc.* **2020**, *142*, 10279−10283.
(408) de Meijere, A.; Kozhushkov, S. I.; Schill, H. Three-Membered-Ring-Based Molecular Architectures. *Chem. Rev.* **2006**, *106*, 4926−4996.
(409) Marek, I.; Simaan, S.; Masarwa, A. Enantiomerically Enriched Cyclopropene Derivatives: Versatile Building Blocks in Asymmetric Synthesis. *Angew. Chem., Int. Ed.* **2007**, *46*, 7364−7376.
(410) Panish, R.; Chintala, S. R.; Boruta, D. T.; Fang, Y.; Taylor, M. T.; Fox, J. M. Enantioselective Synthesis of Cyclobutanes Via Sequential Rh-Catalyzed Bicyclobutanation/Cu-Catalyzed Homoconjugate Addition. *J. Am. Chem. Soc.* **2013**, *135*, 9283−9286.
(411) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. Enzymatic Construction of Highly Strained Carbocycles. *Science* **2018**, *360*, 71−75.
(412) Chen, K.; Arnold, F. H. Engineering Cytochrome P450s for Enantioselective Cyclopropenation of Internal Alkynes. *J. Am. Chem. Soc.* **2020**, *142*, 6891−6895.
(413) Yang, Y.; Cho, I.; Qi, X.; Liu, P.; Arnold, F. H. An Enzymatic Platform for the Asymmetric Amination of Primary, Secondary and Tertiary C(sp^3)−H Bonds. *Nat. Chem.* **2019**, *11*, 987−993.
(414) Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem., Int. Ed.* **2014**, *53*, 6810−6813.
(415) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of De Novo Protein Design. *Nature* **2016**, *537*, 320−327.
(416) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725.
(417) Kries, H.; Blomberg, R.; Hilvert, D. De Novo Enzymes by Computational Design. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221−228.
(418) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190−195.
(419) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Gruetter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision Is Essential for Efficient Catalysis in an Evolved Kemp Eliminase. *Nature* **2013**, *503*, 418−421.
(420) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St. Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; et al. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309−313.
(421) Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Players, F.; Stoddard, B. L.; Popovic, Z.; Baker, D. Increased Diels-Alderase Activity through Backbone Remodeling Guided by Foldit Players. *Nat. Biotechnol.* **2012**, *30*, 190−192.
(422) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovník, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. Impact of Scaffold Rigidity on the Design and Evolution of an Artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8013−8018.
(423) Basler, S.; Studer, S.; Zou, Y.; Mori, T.; Ota, Y.; Camus, A.; Bunzel, H. A.; Helgeson, R. C.; Houk, K. N.; Jiménez-Osés, G.; et al. Efficient Lewis Acid Catalysis of an Abiological Reaction in a De Novo Protein Scaffold. *Nat. Chem.* **2021**, *13*, 231−235.
(424) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; et al. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387−1391.
(425) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494−498.
(426) Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A. D.; Hilvert, D. Efficient Laboratory Evolution of Computationally Designed Enzymes with Low Starting Activities Using Fluorescence-Activated Droplet Sorting. *Protein Eng., Des. Sel.* **2016**, *29*, 355−366.
(427) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. Emergence of a Catalytic Tetrad During Evolution of a Highly Active Artificial Aldolase. *Nat. Chem.* **2017**, *9*, 50−56.
(428) Macdonald, D. S.; Garrabou, X.; Klaus, C.; Verez, R.; Mori, T.; Hilvert, D. Engineered Artificial Carboligases Facilitate Regioselective Preparation of Enantioenriched Aldol Adducts. *J. Am. Chem. Soc.* **2020**, *142*, 10250−10254.
(429) Garrabou, X.; Wicky, B. I. M.; Hilvert, D. Fast Knoevenagel Condensations Catalyzed by an Artificial Schiff-Base-Forming Enzyme. *J. Am. Chem. Soc.* **2016**, *138*, 6972−6974.
(430) Garrabou, X.; Beck, T.; Hilvert, D. A Promiscuous De Novo Retro-Aldolase Catalyzes Asymmetric Michael Additions via Schiff Base Intermediates. *Angew. Chem., Int. Ed.* **2015**, *54*, 5609−5612.
(431) Garrabou, X.; Macdonald, D. S.; Hilvert, D. Chemoselective Henry Condensations Catalyzed by Artificial Carboligases. *Chem. - Eur. J.* **2017**, *23*, 6001−6003.
(432) Garrabou, X.; Macdonald, D. S.; Wicky, B. I. M.; Hilvert, D. Stereodivergent Evolution of Artificial Enzymes for the Michael Reaction. *Angew. Chem., Int. Ed.* **2018**, *57*, 5288−5291.
(433) Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **2013**, *3*, 2954−2975.
(434) Thomson, A. J.; Gray, H. B. Bio-Inorganic Chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2*, 155−158.
(435) Chan, K.-H.; Guan, X.; Lo, V. K.-Y.; Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)−Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chem., Int. Ed.* **2014**, *53*, 2982−2987.
(436) Maxwell, J. L.; O'Malley, S.; Brown, K. C.; Kodadek, T. Shape-Selective and Asymmetric Cyclopropanation of Alkenes Catalyzed by Rhodium Porphyrins. *Organometallics* **1992**, *11*, 645−652.
(437) Anding, B. J.; Ellern, A.; Woo, L. K. Olefin Cyclopropanation Catalyzed by Iridium(III) Porphyrin Complexes. *Organometallics* **2012**, *31*, 3628−3635.
(438) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological Catalysis by Artificial Haem Proteins Containing Noble Metals in Place of Iron. *Nature* **2016**, *534*, 534−537.
(439) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science* **2016**, *354*, 102−106.
(440) Key, H. M.; Dydio, P.; Liu, Z.; Rha, J. Y. E.; Nazarenko, A.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. Beyond Iron: Iridium-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Containing P450 Enzymes for Selective Cyclopropanations of Structurally Diverse Alkenes. *ACS Cent. Sci.* **2017**, *3*, 302−308.
(441) Gu, Y.; Natoli, S. N.; Liu, Z.; Clark, D. S.; Hartwig, J. F. Site-Selective Functionalization of (sp^3)C-H Bonds Catalyzed by Artificial Metalloenzymes Containing an Iridium-Porphyrin Cofactor. *Angew. Chem., Int. Ed.* **2019**, *58*, 13954−13960.
(442) Reetz, M. T.; Rentzsch, M.; Pletsch, A.; Taglieber, A.; Hollmann, F.; Mondière, R. J.; Dickmann, N.; Höcker, B.; Cerrone, S.; Haeger, M. C.; et al. A Robust Protein Host for Anchoring Chelating Ligands and Organocatalysts. *ChemBioChem* **2008**, *9*, 552−564.
(443) Köhler, V.; Wilson, Y. M.; Lo, C.; Sardo, A.; Ward, T. R. Protein-Based Hybrid Catalysts—Design and Evolution. *Curr. Opin. Biotechnol.* **2010**, *21*, 744−752.
(444) Nicolas, I.; Le Maux, P.; Simonneaux, G. Asymmetric Catalytic Cyclopropanation Reactions in Water. *Coord. Chem. Rev.* **2008**, *252*, 727−735.
(445) Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C. A General Method for Artificial Metalloenzyme Formation through Strain-Promoted Azide−Alkyne Cycloaddition. *ChemBioChem* **2014**, *15*, 223−227.
(446) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. Engineering a Dirhodium Artificial Metalloenzyme for Selective Olefin Cyclopropanation. *Nat. Commun.* **2015**, *6*, 7789.
(447) Yang, H.; Swartz, A. M.; Park, H. J.; Srivastava, P.; Ellis-Guardiola, K.; Upp, D. M.; Lee, G.; Belsare, K.; Gu, Y.; Zhang, C.; et al. Evolving Artificial Metalloenzymes via Random Mutagenesis. *Nat. Chem.* **2018**, *10*, 318−324.
(448) Heinisch, T.; Ward, T. R. Artificial Metalloenzymes Based on the Biotin−Streptavidin Technology: Challenges and Opportunities. *Acc. Chem. Res.* **2016**, *49*, 1711−1721.
(449) Skander, M.; Humbert, N.; Collot, J.; Gradinaru, J.; Klein, G.; Loosli, A.; Sauser, J.; Zocchi, A.; Gilardoni, F.; Ward, T. R. Artificial Metalloenzymes: (Strept)Avidin as Host for Enantioselective Hydrogenation by Achiral Biotinylated Rhodium-Diphosphine Complexes. *J. Am. Chem. Soc.* **2004**, *126*, 14411−14418.
(450) Klein, G.; Humbert, N.; Gradinaru, J.; Ivanova, A.; Gilardoni, F.; Rusbandi, U. E.; Ward, T. R. Tailoring the Active Site of Chemzymes by Using a Chemogenetic-Optimization Procedure: Towards Substrate-Specific Artificial Hydrogenases Based on the Biotin−Avidin Technology. *Angew. Chem., Int. Ed.* **2005**, *44*, 7764−7767.
(451) Letondor, C.; Humbert, N.; Ward, T. R. Artificial Metalloenzymes Based on Biotin-Avidin Technology for the Enantioselective Reduction of Ketones by Transfer Hydrogenation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 4683.
(452) Creus, M.; Pordea, A.; Rossel, T.; Sardo, A.; Letondor, C.; Ivanova, A.; LeTrong, I.; Stenkamp, R. E.; Ward, T. R. X-Ray Structure and Designed Evolution of an Artificial Transfer Hydrogenase. *Angew. Chem., Int. Ed.* **2008**, *47*, 1400−1404.
(453) Dürrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knörr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; et al. Artificial Transfer Hydrogenases for the Enantioselective Reduction of Cyclic Imines. *Angew. Chem., Int. Ed.* **2011**, *50*, 3026−3029.
(454) Zimbron, J. M.; Heinisch, T.; Schmid, M.; Hamels, D.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. A Dual Anchoring Strategy for the Localization and Activation of Artificial Metalloenzymes Based on the Biotin−Streptavidin Technology. *J. Am. Chem. Soc.* **2013**, *135*, 5384−5388.
(455) Pierron, J.; Malan, C.; Creus, M.; Gradinaru, J.; Hafner, I.; Ivanova, A.; Sardo, A.; Ward, T. R. Artificial Metalloenzymes for Asymmetric Allylic Alkylation on the Basis of the Biotin−Avidin Technology. *Angew. Chem., Int. Ed.* **2008**, *47*, 701−705.
(456) Köhler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knörr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; et al. OsO4·Streptavidin: A Tunable Hybrid Catalyst for the Enantioselective cis-Dihydroxylation of Olefins. *Angew. Chem., Int. Ed.* **2011**, *50*, 10863−10866.
(457) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. Artificial Metalloenzyme for Enantioselective Sulfoxidation Based on Vanadyl-Loaded Streptavidin. *J. Am. Chem. Soc.* **2008**, *130*, 8085−8088.
(458) Thomas, C. M.; Letondor, C.; Humbert, N.; Ward, T. R. Aqueous Oxidation of Alcohols Catalyzed by Artificial Metalloenzymes Based on the Biotin−Avidin Technology. *J. Organomet. Chem.* **2005**, *690*, 4488−4491.
(459) Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R. An Enantioselective Artificial Suzukiase Based on the Biotin-Streptavidin Technology. *Chem. Sci.* **2016**, *7*, 673−677.
(460) Hyster, T. K.; Knörr, L.; Ward, T. R.; Rovis, T. Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C−H Activation. *Science* **2012**, *338*, 500−503.
(461) Lo, C.; Ringenberg, M. R.; Gnandt, D.; Wilson, Y.; Ward, T. R. Artificial Metalloenzymes for Olefin Metathesis Based on the Biotin-(Strept)Avidin Technology. *Chem. Commun.* **2011**, *47*, 12065−12067.
(462) Mallin, H.; Hestericová, M.; Reuter, R.; Ward, T. R. Library Design and Screening Protocol for Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology. *Nat. Protoc.* **2016**, *11*, 835−852.
(463) Wilson, Y. M.; Dürrenberger, M.; Nogueira, E. S.; Ward, T. R. Neutralizing the Detrimental Effect of Glutathione on Precious Metal Catalysts. *J. Am. Chem. Soc.* **2014**, *136*, 8928−8932.
(464) Hestericová, M.; Heinisch, T.; Alonso-Cotchico, L.; Maréchal, J.-D.; Vidossich, P.; Ward, T. R. Directed Evolution of an Artificial Imine Reductase. *Angew. Chem., Int. Ed.* **2018**, *57*, 1863−1868.
(465) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. Directed Evolution of Artificial Metalloenzymes for in Vivo Metathesis. *Nature* **2016**, *537*, 661−665.
(466) Okamoto, Y.; Kojima, R.; Schwizer, F.; Bartolami, E.; Heinisch, T.; Matile, S.; Fussenegger, M.; Ward, T. R. A Cell-Penetrating Artificial Metalloenzyme Regulates a Gene Switch in a Designer Mammalian Cell. *Nat. Commun.* **2018**, *9*, 1943.
(467) Drienovská, I.; Roelfes, G. Expanding the Enzyme Universe with Genetically Encoded Unnatural Amino Acids. *Nat. Catal.* **2020**, *3*, 193−202.
(468) Pott, M.; Hayashi, T.; Mori, T.; Mittl, P. R. E.; Green, A. P.; Hilvert, D. A Noncanonical Proximal Heme Ligand Affords an Efficient Peroxidase in a Globin Fold. *J. Am. Chem. Soc.* **2018**, *140*, 1535−1543.
(469) Mayer, C.; Dulson, C.; Reddem, E.; Thunnissen, A.-M. W. H.; Roelfes, G. Directed Evolution of a Designer Enzyme Featuring an Unnatural Catalytic Amino Acid. *Angew. Chem., Int. Ed.* **2019**, *58*, 2083−2087.
(470) Burke, A. J.; Lovelock, S. L.; Frese, A.; Crawshaw, R.; Ortmayer, M.; Dunstan, M.; Levy, C.; Green, A. P. Design and Evolution of an Enzyme with a Non-Canonical Organocatalytic Mechanism. *Nature* **2019**, *570*, 219−223.
(471) Xiao, H.; Nasertorabi, F.; Choi, S.-h.; Han, G. W.; Reed, S. A.; Stevens, R. C.; Schultz, P. G. Exploring the Potential Impact of an Expanded Genetic Code on Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6961−6966.
(472) Carroll, D. Genome Engineering with Targetable Nucleases. *Annu. Rev. Biochem.* **2014**, *83*, 409−439.
(473) Gaj, T.; Gersbach, C. A.; Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering. *Trends Biotechnol.* **2013**, *31*, 397−405.
(474) Klug, A. The Discovery of Zinc Fingers and Their Development for Practical Applications in Gene Regulation and Genome Manipulation. *Q. Rev. Biophys.* **2010**, *43*, 1−21.
(475) Maeder, M. L.; Thibodeau-Beganny, S.; Osiak, A.; Wright, D. A.; Anthony, R. M.; Eichtinger, M.; Jiang, T.; Foley, J. E.; Winfrey, R. J.; Townsend, J. A.; et al. Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Mol. Cell* **2008**, *31*, 294−301.
(476) Beerli, R. R.; Segal, D. J.; Dreier, B.; Barbas, C. F. Toward Controlling Gene Expression at Will: Specific Regulation of the erbB-

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

2/HER-2 Promoter by Using Polydactyl Zinc Finger Proteins Constructed from Modular Building Blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14628−14633.
(477) Guo, J.; Gaj, T.; Barbas, C. F., 3rd Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *J. Mol. Biol.* **2010**, *400*, 96−107.
(478) Moscou, M. J.; Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **2009**, *326*, 1501−1501.
(479) Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **2009**, *326*, 1509−1512.
(480) Deng, D.; Yan, C.; Pan, X.; Mahfouz, M.; Wang, J.; Zhu, J.-K.; Shi, Y.; Yan, N. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science* **2012**, *335*, 720−723.
(481) Mak, A. N.-S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **2012**, *335*, 716−719.
(482) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced Proofreading Governs CRISPR−Cas9 Targeting Accuracy. *Nature* **2017**, *550*, 407−410.
(483) Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P. W.; Li, Z.; Peterson, R. T.; Yeh, J.-R. J.; et al. Engineered CRISPR-Cas9 Nucleases with Altered Pam Specificities. *Nature* **2015**, *523*, 481−485.
(484) Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K. High-Fidelity CRISPR−Cas9 Nucleases with No Detectable Genome-Wide Off-Target Effects. *Nature* **2016**, *529*, 490−495.
(485) Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X.; Zhang, F. Rationally Engineered Cas9 Nucleases with Improved Specificity. *Science* **2016**, *351*, 84−88.
(486) Kim, S.; Bae, T.; Hwang, J.; Kim, J.-S. Rescue of High-Specificity Cas9 Variants Using sgRNAs with Matched 5′ Nucleotides. *Genome Biol.* **2017**, *18*, 218.
(487) Kulcsár, P. I.; Tálas, A.; Huszár, K.; Ligeti, Z.; Tóth, E.; Weinhardt, N.; Fodor, E.; Welker, E. Crossing Enhanced and High Fidelity SpCas9 Nucleases to Optimize Specificity and Cleavage. *Genome Biol.* **2017**, *18*, 190.
(488) Zhang, D.; Zhang, H.; Li, T.; Chen, K.; Qiu, J.-L.; Gao, C. Perfectly Matched 20-Nucleotide Guide RNA Sequences Enable Robust Genome Editing Using High-Fidelity SpCas9 Nucleases. *Genome Biol.* **2017**, *18*, 191.
(489) Anderson, K. R.; Haeussler, M.; Watanabe, C.; Janakiraman, V.; Lund, J.; Modrusan, Z.; Stinson, J.; Bei, Q.; Buechler, A.; Yu, C.; et al. CRISPR Off-Target Analysis in Genetically Engineered Rats and Mice. *Nat. Methods* **2018**, *15*, 512−514.
(490) Casini, A.; Olivieri, M.; Petris, G.; Montagna, C.; Reginato, G.; Maule, G.; Lorenzin, F.; Prandi, D.; Romanel, A.; Demichelis, F.; et al. A Highly Specific SpCas9 Variant Is Identified by in Vivo Screening in Yeast. *Nat. Biotechnol.* **2018**, *36*, 265−271.
(491) Lee, J. K.; Jeong, E.; Lee, J.; Jung, M.; Shin, E.; Kim, Y.-h.; Lee, K.; Jung, I.; Kim, D.; Kim, S.; et al. Directed Evolution of CRISPR-Cas9 to Increase Its Specificity. *Nat. Commun.* **2018**, *9*, 3048.
(492) Thuronyi, B. W.; Koblan, L. W.; Levy, J. M.; Yeh, W.-H.; Zheng, C.; Newby, G. A.; Wilson, C.; Bhaumik, M.; Shubina-Oleinik, O.; Holt, J. R.; et al. Continuous Evolution of Base Editors with Expanded Target Compatibility and Improved Activity. *Nat. Biotechnol.* **2019**, *37*, 1070−1079.
(493) Dunn, H. A.; Orlandi, C.; Martemyanov, K. A. Beyond the Ligand: Extracellular and Transcellular G Protein-Coupled Receptor Complexes in Physiology and Pharmacology. *Pharmacol. Rev.* **2019**, *71*, 503−519.
(494) Sriram, K.; Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **2018**, *93*, 251−258.
(495) Jo, M.; Jung, S. T. Engineering Therapeutic Antibodies Targeting G-Protein-Coupled Receptors. *Exp. Mol. Med.* **2016**, *48*, No. e207.
(496) Lundstrom, K. Structural Genomics of GPCRs. *Trends Biotechnol.* **2005**, *23*, 103−108.
(497) Sarramegna, V.; Talmont, F.; Demange, P.; Milon, A. Heterologous Expression of G-Protein-Coupled Receptors: Comparison of Expression Systems from the Standpointof Large-Scale Production and Purification. *Cell. Mol. Life Sci.* **2003**, *60*, 1529−1546.
(498) Sarkar, C. A.; Dodevski, I.; Kenig, M.; Dudli, S.; Mohr, A.; Hermans, E.; Pluckthun, A. Directed Evolution of a G Protein-Coupled Receptor for Expression, Stability, and Binding Selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14808−14813.
(499) Dodevski, I.; Plückthun, A. Evolution of Three Human GPCRs for Higher Expression and Stability. *J. Mol. Biol.* **2011**, *408*, 599−615.
(500) Schlinkmann, K. M.; Plückthun, A. Directed Evolution of G-Protein-Coupled Receptors for High Functional Expression and Detergent Stability. *Methods Enzymol.* **2013**, *520*, 67−97.
(501) Waltenspühl, Y.; Jeliazkov, J. R.; Kummer, L.; Plückthun, A. Directed Evolution for High Functional Production and Stability of a Challenging G Protein-Coupled Receptor. *Sci. Rep.* **2021**, *11*, 8630.
(502) Krishnamurthy, A.; Jimeno, A. Bispecific Antibodies for Cancer Therapy: A Review. *Pharmacol. Ther.* **2018**, *185*, 122−134.
(503) Hooks, M. A.; Wade, C. S.; Millikan, W. J., Jr. Muromonab CD-3: A Review of Its Pharmacology, Pharmacokinetics, and Clinical Use in Transplantation. *Pharmacotherapy* **1991**, *11*, 26−37.
(504) Singh, S.; Kumar, N. K.; Dwiwedi, P.; Charan, J.; Kaur, R.; Sidhu, P.; Chugh, V. K. Monoclonal Antibodies: A Review. *Curr. Clin. Pharmacol.* **2018**, *13*, 85−99.
(505) Payne, W. J., Jr; Marshall, D. L.; Shockley, R. K.; Martin, W. J. Clinical Laboratory Applications of Monoclonal Antibodies. *Clin. Microbiol. Rev.* **1988**, *1*, 313−329.
(506) Ministro, J.; Manuel, A. M.; Goncalves, J. In *Current Applications of Pharmaceutical Biotechnology*; Silva, A. C., Moreira, J. N., Lobo, J. M. S., Almeida, H., Eds.; Springer International Publishing: Cham, 2020.
(507) Boder, E. T.; Midelfort, K. S.; Wittrup, K. D. Directed Evolution of Antibody Fragments with Monovalent Femtomolar Antigen-Binding Affinity. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10701−10705.
(508) Xu, L.; Aha, P.; Gu, K.; Kuimelis, R. G.; Kurz, M.; Lam, T.; Lim, A. C.; Liu, H.; Lohse, P. A.; Sun, L.; et al. Directed Evolution of High-Affinity Antibody Mimics Using mRNA Display. *Chem. Biol.* **2002**, *9*, 933−942.
(509) Jermutus, L.; Honegger, A.; Schwesinger, F.; Hanes, J.; Plückthun, A. Tailoring in Vitro Evolution for Protein Affinity or Stability. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 75−80.
(510) Zahnd, C.; Spinelli, S.; Luginbühl, B.; Amstutz, P.; Cambillau, C.; Plückthun, A. Directed in Vitro Evolution and Crystallographic Analysis of a Peptide-Binding Single Chain Antibody Fragment (scFv) with Low Picomolar Affinity. *J. Biol. Chem.* **2004**, *279*, 18870−18877.
(511) Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K. Adrenaline-Activated Structure of B2-Adrenoceptor Stabilized by an Engineered Nanobody. *Nature* **2013**, *502*, 575−579.
(512) Sadelain, M.; Rivière, I.; Riddell, S. Therapeutic T Cell Engineering. *Nature* **2017**, *545*, 423−431.
(513) Shusta, E. V.; Holler, P. D.; Kieke, M. C.; Kranz, D. M.; Wittrup, K. D. Directed Evolution of a Stable Scaffold for T-Cell Receptor Engineering. *Nat. Biotechnol.* **2000**, *18*, 754−759.
(514) Li, Y.; Moysey, R.; Molloy, P. E.; Vuidepot, A.-L.; Mahon, T.; Baston, E.; Dunn, S.; Liddy, N.; Jacob, J.; Jakobsen, B. K.; et al. Directed Evolution of Human T-Cell Receptors with Picomolar Affinities by Phage Display. *Nat. Biotechnol.* **2005**, *23*, 349−354.
(515) Li, J.-W.; Zhang, X.-Y.; Wu, H.; Bai, Y.-P. Transcription Factor Engineering for High-Throughput Strain Evolution and Organic Acid Bioproduction: A Review. *Front. Bioeng. Biotechnol.* **2020**, *8*, 8.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(516) Collins, C. H.; Leadbetter, J. R.; Arnold, F. H. Dual Selection Enhances the Signaling Specificity of a Variant of the Quorum-Sensing Transcriptional Activator LuxR. *Nat. Biotechnol.* **2006**, *24*, 708−712.
(517) Tang, S.-Y.; Fazelinia, H.; Cirino, P. C. AraC Regulatory Protein Mutants with Altered Effector Specificity. *J. Am. Chem. Soc.* **2008**, *130*, 5267−5271.
(518) Ellefson, J. W.; Ledbetter, M. P.; Ellington, A. D. Directed Evolution of a Synthetic Phylogeny of Programmable Trp Repressors. *Nat. Chem. Biol.* **2018**, *14*, 361−367.
(519) Seok, J. Y.; Yang, J.; Choi, S. J.; Lim, H. G.; Choi, U. J.; Kim, K.-J.; Park, S.; Yoo, T. H.; Jung, G. Y. Directed Evolution of the 3-Hydroxypropionic Acid Production Pathway by Engineering Aldehyde Dehydrogenase Using a Synthetic Selection Device. *Metab. Eng.* **2018**, *47*, 113−120.
(520) Kell, D. B.; Swainston, N.; Pir, P.; Oliver, S. G. Membrane Transporter Engineering in Industrial Biotechnology and Whole Cell Biocatalysis. *Trends Biotechnol.* **2015**, *33*, 237−246.
(521) Lian, J.; Li, Y.; HamediRad, M.; Zhao, H. Directed Evolution of a Cellodextrin Transporter for Improved Biofuel Production under Anaerobic Conditions in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **2014**, *111*, 1521−1531.
(522) Zhu, Z.; Hu, Y.; Teixeira, P. G.; Pereira, R.; Chen, Y.; Siewers, V.; Nielsen, J. Multidimensional Engineering of *Saccharomyces cerevisiae* for Efficient Synthesis of Medium-Chain Fatty Acids. *Nat. Catal.* **2020**, *3*, 64−74.
(523) Foo, J. L.; Leong, S. S. J. Directed Evolution of an *E. coli* Inner Membrane Transporter for Improved Efflux of Biofuel Molecules. *Biotechnol. Biofuels* **2013**, *6*, 81.
(524) Fisher, M. A.; Boyarskiy, S.; Yamada, M. R.; Kong, N.; Bauer, S.; Tullman-Ercek, D. Enhancing Tolerance to Short-Chain Alcohols by Engineering the *Escherichia coli* Acrb Efflux Pump to Secrete the Non-Native Substrate N-Butanol. *ACS Synth. Biol.* **2014**, *3*, 30−40.
(525) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.
(526) Sawano, A.; Miyawaki, A. Directed Evolution of Green Fluorescent Protein by a New Versatile PCR Strategy for Site-Directed and Semi-Random Mutagenesis. *Nucleic Acids Res.* **2000**, *28*, No. E78.
(527) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene* **1996**, *173*, 33−38.
(528) Volkov, A. A.; Shao, Z.; Arnold, F. H. Recombination and Chimeragenesis by in Vitro Heteroduplex Formation and in Vivo Repair. *Nucleic Acids Res.* **1999**, *27*, No. e18.
(529) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 12501−12504.
(530) Yang, T.-T.; Cheng, L.; Kain, S. R. Optimized Codon Usage and Chromophore Mutations Provide Enhanced Sensitivity with the Green Fluorescent Protein. *Nucleic Acids Res.* **1996**, *24*, 4592−4593.
(531) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. A Monomeric Red Fluorescent Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877−7882.
(532) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. Biochemistry, Mutagenesis, and Oligomerization of DsRed, a Red Fluorescent Protein from Coral. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11984−11989.
(533) Bevis, B. J.; Glick, B. S. Rapidly Maturing Variants of the Discosoma Red Fluorescent Protein (DsRed). *Nat. Biotechnol.* **2002**, *20*, 83−87.
(534) Gurskaya, N. G.; Fradkov, A. F.; Terskikh, A.; Matz, M. V.; Labas, Y. A.; Martynov, V. I.; Yanushevich, Y. G.; Lukyanov, K. A.; Lukyanov, S. A. GFP-Like Chromoproteins as a Source of Far-Red Fluorescent Proteins. *FEBS Lett.* **2001**, *507*, 16−20.
(535) Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A. A Variant of Yellow Fluorescent Protein with Fast and Efficient Maturation for Cell-Biological Applications. *Nat. Biotechnol.* **2002**, *20*, 87−90.
(536) Du, J.; Yuan, Y.; Si, T.; Lian, J.; Zhao, H. Customized Optimization of Metabolic Pathways by Combinatorial Transcriptional Engineering. *Nucleic Acids Res.* **2012**, *40*, No. e142.
(537) Yuan, Y.; Du, J.; Zhao, H. Customized Optimization of Metabolic Pathways by Combinatorial Transcriptional Engineering. *Methods Mol. Biol.* **2013**, *985*, 177.
(538) Kim, B.; Du, J.; Eriksen, D. T.; Zhao, H. Combinatorial Design of a Highly Efficient Xylose-Utilizing Pathway in *Saccharomyces cerevisiae* for the Production of Cellulosic Biofuels. *Appl. Environ. Microbiol.* **2013**, *79*, 931.
(539) Latimer, L. N.; Lee, M. E.; Medina-Cleghorn, D.; Kohnz, R. A.; Nomura, D. K.; Dueber, J. E. Employing a Combinatorial Expression Approach to Characterize Xylose Utilization in *Saccharomyces cerevisiae*. *Metab. Eng.* **2014**, *25*, 20−29.
(540) Eriksen, D. T.; Hsieh, P. C. H.; Lynn, P.; Zhao, H. Directed Evolution of a Cellobiose Utilization Pathway in *Saccharomyces cerevisiae* by Simultaneously Engineering Multiple Proteins. *Microb. Cell Fact.* **2013**, *12*, 61.
(541) Yuan, Y.; Zhao, H. Directed Evolution of a Highly Efficient Cellobiose Utilizing Pathway in an Industrial *Saccharomyces cerevisiae* Strain. *Biotechnol. Bioeng.* **2013**, *110*, 2874−2881.
(542) Shen, Y. P.; Fong, L. S.; Yan, Z. B.; Liu, J. Z. Combining Directed Evolution of Pathway Enzymes and Dynamic Pathway Regulation Using a Quorum-Sensing Circuit to Improve the Production of 4-Hydroxyphenylacetic Acid in *Escherichia coli*. *Biotechnol. Biofuels* **2019**, *12*, 94.
(543) Raman, S.; Rogers, J. K.; Taylor, N. D.; Church, G. M. Evolution-Guided Optimization of Biosynthetic Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17803−17808.
(544) Leavitt, J. M.; Wagner, J. M.; Tu, C. C.; Tong, A.; Liu, Y.; Alper, H. S. Biosensor-Enabled Directed Evolution to Improve Muconic Acid Production in *Saccharomyces cerevisiae*. *Biotechnol. J.* **2017**, *12*, 1600687.
(545) Schmidt-Dannert, C.; Umeno, D.; Arnold, F. H. Molecular Breeding of Carotenoid Biosynthetic Pathways. *Nat. Biotechnol.* **2000**, *18*, 750.
(546) Chen, Y.-Y.; Shen, H.-J.; Cui, Y.-Y.; Chen, S.-G.; Weng, Z.-M.; Zhao, M.; Liu, J.-Z. Chromosomal Evolution of *Escherichia coli* for the Efficient Production of Lycopene. *BMC Biotechnol.* **2013**, *13*, 6.
(547) Martin, V. J. J.; Piteral, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. Engineering a Mevalonate Pathway in *Escherichia coli* for Production of Terpenoids. *Nat. Biotechnol.* **2003**, *21*, 796−802.
(548) Pfleger, B. F.; Pitera, D. J.; Smolke, C. D.; Keasling, J. D. Combinatorial Engineering of Intergenic Regions in Operons Tunes Expression of Multiple Genes. *Nat. Biotechnol.* **2006**, *24*, 1027−1032.
(549) Ng, C. Y.; Farasat, I.; Maranas, C. D.; Salis, H. M. Rational Design of a Synthetic Entner-Doudoroff Pathway for Improved and Controllable Nadph Regeneration. *Metab. Eng.* **2015**, *29*, 86−96.
(550) Wang, H. H.; Kim, H.; Cong, L.; Jeong, J.; Bang, D.; Church, G. M. Genome-Scale Promoter Engineering by Coselection MAGE. *Nat. Methods* **2012**, *9*, 591−593.
(551) Yokobayashi, Y.; Weiss, R.; Arnold, F. H. Directed Evolution of a Genetic Circuit. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16587−16591.
(552) Anderson, J. C.; Clarke, E. J.; Arkin, A. P.; Voigt, C. A. Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria. *J. Mol. Biol.* **2006**, *355*, 619−627.
(553) Yokobayashi, Y.; Arnold, F. H. A Dual Selection Module for Directed Evolution of Genetic Circuits. *Nat. Comput.* **2005**, *4*, 245−254.
(554) Nomura, Y.; Yokobayashi, Y. Dual Selection of a Genetic Switch by a Single Selection Marker. *BioSystems* **2007**, *90*, 115−120.
(555) Tashiro, Y.; Fukutomi, H.; Terakubo, K.; Saito, K.; Umeno, D. A Nucleoside Kinase as a Dual Selector for Genetic Switches and Circuits. *Nucleic Acids Res.* **2011**, *39*, No. e12.
(556) Kimura, Y.; Kawai-Noma, S.; Saito, K.; Umeno, D. Directed Evolution of the Stringency of the LuxR Vibrio Fischeri Quorum

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

Sensor without OFF-State Selection. *ACS Synth. Biol.* **2020**, *9*, 567−575.
(557) Ellefson, J. W.; Meyer, A. J.; Hughes, R. A.; Cannon, J. R.; Brodbelt, J. S.; Ellington, A. D. Directed Evolution of Genetic Parts and Circuits by Compartmentalized Partnered Replication. *Nat. Biotechnol.* **2014**, *32*, 97−101.
(558) Abil, Z.; Ellefson, J. W.; Gollihar, J. D.; Watkins, E.; Ellington, A. D. Compartmentalized Partnered Replication for the Directed Evolution of Genetic Parts and Circuits. *Nat. Protoc.* **2017**, *12*, 2493−2512.
(559) Schaerli, Y.; Isalan, M. Building Synthetic Gene Circuits from Combinatorial Libraries: Screening and Selection Strategies. *Mol. BioSyst.* **2013**, *9*, 1559−1567.
(560) Liu, C.; Fu, X.; Liu, L.; Ren, X.; Chau, C. K.; Li, S.; Xiang, L.; Zeng, H.; Chen, G.; Tang, L.-H. Sequential Establishment of Stripe Patterns in an Expanding Cell Population. *Science* **2011**, *334*, 238−241.
(561) Guantes, R.; Poyatos, J. F. Dynamical Principles of Two-Component Genetic Oscillators. *PLoS Comput. Biol.* **2006**, *2*, No. e30.
(562) Brophy, J. A. N.; Voigt, C. A. Principles of Genetic Circuit Design. *Nat. Methods* **2014**, *11*, 508−520.
(563) Chuang, J. S. Engineering Multicellular Traits in Synthetic Microbial Populations. *Curr. Opin. Chem. Biol.* **2012**, *16*, 370−378.
(564) Sanjuán, R.; Domingo-Calap, P. Mechanisms of Viral Mutation. *Cell. Mol. Life Sci.* **2016**, *73*, 4433−4448.
(565) Duffy, S.; Shackelton, L. A.; Holmes, E. C. Rates of Evolutionary Change in Viruses: Patterns and Determinants. *Nat. Rev. Genet.* **2008**, *9*, 267−276.
(566) Plant, E. P.; Manukyan, H.; Laassri, M.; Ye, Z. Insights from the Comparison of Genomic Variants from Two Influenza B Viruses Grown in the Presence of Human Antibodies in Cell Culture. *PLoS One* **2020**, *15*, No. e0239015.
(567) Villarreal, L. P. In *Origin and Evolution of Viruses*; Domingo, E., Webster, R., Holland, J., Eds.; Academic Press: London, 1999.
(568) Koonin, E. V.; Senkevich, T. G.; Dolja, V. V. The Ancient Virus World and Evolution of Cells. *Biol. Direct* **2006**, *1*, 29.
(569) Moelling, K.; Broecker, F. Viruses and Evolution − Viruses First? A Personal Perspective. *Front. Microbiol.* **2019**, *10*, 10.
(570) Elena, S. F.; Sanjuán, R. Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 27−52.
(571) Dolan, P. T.; Whitfield, Z. J.; Andino, R. Mapping the Evolutionary Potential of RNA Viruses. *Cell Host Microbe* **2018**, *23*, 435−446.
(572) Das, A. T.; Zhou, X.; Vink, M.; Klaver, B.; Verhoef, K.; Marzio, G.; Berkhout, B. Viral Evolution as a Tool to Improve the Tetracycline-Regulated Gene Expression System. *J. Biol. Chem.* **2004**, *279*, 18776−18782.
(573) Makeyev, E. V.; Bamford, D. H. Evolutionary Potential of an RNA Virus. *J. Virol.* **2004**, *78*, 2114−2120.
(574) Na, D.; Lee, S.; Yi, G. S.; Lee, D. Synthetic Inter-Species Cooperation of Host and Virus for Targeted Genetic Evolution. *J. Biotechnol.* **2011**, *153*, 35−41.
(575) Wang, D.; Tai, P. W. L.; Gao, G. Adeno-Associated Virus Vector as a Platform for Gene Therapy Delivery. *Nat. Rev. Drug Discovery* **2019**, *18*, 358−378.
(576) Bartel, M. A.; Weinstein, J. R.; Schaffer, D. V. Directed Evolution of Novel Adeno-Associated Viruses for Therapeutic Gene Delivery. *Gene Ther.* **2012**, *19*, 694−700.
(577) Grimm, D.; Kay, M. A. From Virus Evolution to Vector Revolution: Use of Naturally Occurring Serotypes of Adeno-Associated Virus (AAV) as Novel Vectors for Human Gene Therapy. *Curr. Gene Ther.* **2003**, *3*, 281−304.
(578) El Andari, J.; Grimm, D. Production, Processing, and Characterization of Synthetic AAV Gene Therapy Vectors. *Biotechnol. J.* **2021**, *16*, 2000025.
(579) Lisowski, L.; Tay, S. S.; Alexander, I. E. Adeno-Associated Virus Serotypes for Gene Therapeutics. *Curr. Opin. Pharmacol.* **2015**, *24*, 59−67.
(580) Maheshri, N.; Koerber, J. T.; Kaspar, B. K.; Schaffer, D. V. Directed Evolution of Adeno-Associated Virus Yields Enhanced Gene Delivery Vectors. *Nat. Biotechnol.* **2006**, *24*, 198−204.
(581) Wooley, D. P.; Sharma, P.; Weinstein, J. R.; Kotha Lakshmi Narayan, P.; Schaffer, D. V.; Excoffon, K. A Directed Evolution Approach to Select for Novel Adeno-Associated Virus Capsids on an Hiv-1 Producer T Cell Line. *J. Virol. Methods* **2017**, *250*, 47−54.
(582) Dalkara, D.; Byrne, L. C.; Klimczak, R. R.; Visel, M.; Yin, L.; Merigan, W. H.; Flannery, J. G.; Schaffer, D. V. In Vivo-Directed Evolution of a New Adeno-Associated Virus for Therapeutic Outer Retinal Gene Delivery from the Vitreous. *Sci. Transl. Med.* **2013**, *5*, 189ra76.
(583) Byrne, L. C.; Day, T. P.; Visel, M.; Strazzeri, J. A.; Fortuny, C.; Dalkara, D.; Merigan, W. H.; Schaffer, D. V.; Flannery, J. G. In Vivo-Directed Evolution of Adeno-Associated Virus in the Primate Retina. *JCI Insight* **2020**, *5*, 5.
(584) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; et al. Evolution of a Designed Protein Assembly Encapsulating Its Own RNA Genome. *Nature* **2017**, *552*, 415−420.
(585) Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory Evolution of Virus-Like Nucleocapsids from Nonviral Protein Cages. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 5432.
(586) Cao, M.; Tran, V. G.; Zhao, H. Unlocking Nature's Biosynthetic Potential by Directed Genome Evolution. *Curr. Opin. Biotechnol.* **2020**, *66*, 95−104.
(587) Zhang, W.; Geng, A. Improved Ethanol Production by a Xylose-Fermenting Recombinant Yeast Strain Constructed through a Modified Genome Shuffling Method. *Biotechnol. Biofuels* **2012**, *5*, 46.
(588) HamediRad, M.; Lian, J.; Li, H.; Zhao, H. RNAi Assisted Genome Evolution Unveils Yeast Mutants with Improved Xylose Utilization. *Biotechnol. Bioeng.* **2018**, *115*, 1552−1560.
(589) Pereira, F.; Azevedo, F.; Parachin, N. S.; Hahn-Hägerdal, B.; Gorwa-Grauslund, M. F.; Johansson, B. Yeast Pathway Kit: A Method for Metabolic Pathway Assembly with Automatically Simulated Executable Documentation. *ACS Synth. Biol.* **2016**, *5*, 386−394.
(590) Otte, B.; Grunwaldt, E.; Mahmoud, O.; Jennewein, S. Genome Shuffling in *Clostridium Diolis* DSM 15410 for Improved 1,3-Propanediol Production. *Appl. Environ. Microbiol.* **2009**, *75*, 7610.
(591) Si, T.; Chao, R.; Min, Y.; Wu, Y.; Ren, W.; Zhao, H. Automated Multiplex Genome-Scale Engineering in Yeast. *Nat. Commun.* **2017**, *8*, 15187.
(592) Lee, K.-S.; Hong, M.-E.; Jung, S.-C.; Ha, S.-J.; Yu, B. J.; Koo, H. M.; Park, S. M.; Seo, J.-H.; Kweon, D.-H.; Park, J. C.; et al. Improved Galactose Fermentation of *Saccharomyces cerevisiae* through Inverse Metabolic Engineering. *Biotechnol. Bioeng.* **2011**, *108*, 621−631.
(593) Pinel, D.; D'Aoust, F.; del Cardayre, S. B.; Bajwa, P. K.; Lee, H.; Martin, V. J. J. *Saccharomyces cerevisiae* Genome Shuffling through Recursive Population Mating Leads to Improved Tolerance to Spent Sulfite Liquor. *Appl. Environ. Microbiol.* **2011**, *77*, 4736−4743.
(594) Sandberg, T. E.; Salazar, M. J.; Weng, L. L.; Palsson, B. O.; Feist, A. M. The Emergence of Adaptive Laboratory Evolution as an Efficient Tool for Biological Discovery and Industrial Biotechnology. *Metab. Eng.* **2019**, *56*, 1−16.
(595) Xiao, H.; Zhao, H. Genome-Wide RNAi Screen Reveals the E3 SUMO-Protein Ligase Gene *SIZ1* as a Novel Determinant of Furfural Tolerance in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **2014**, *7*, 78.
(596) Wang, W.; Wu, B.; Qin, H.; Liu, P.; Qin, Y.; Duan, G.; Hu, G.; He, M. Genome Shuffling Enhances Stress Tolerance of *Zymomonas Mobilis* to Two Inhibitors. *Biotechnol. Biofuels* **2019**, *12*, 288−288.
(597) Si, T.; Luo, Y.; Bao, Z.; Zhao, H. RNAi-Assisted Genome Evolution in *Saccharomyces cerevisiae* for Complex Phenotype Engineering. *ACS Synth. Biol.* **2015**, *4*, 283−291.
(598) Crook, N.; Sun, J.; Morse, N.; Schmitz, A.; Alper, H. S. Identification of Gene Knockdown Targets Conferring Enhanced Isobutanol and 1-Butanol Tolerance to *Saccharomyces cerevisiae* Using

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

a Tunable RNAi Screening Approach. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 10005−10018.
(599) Warner, J. R.; Reeder, P. J.; Karimpour-Fard, A.; Woodruff, L. B. A.; Gill, R. T. Rapid Profiling of a Microbial Genome Using Mixtures of Barcoded Oligonucleotides. *Nat. Biotechnol.* **2010**, *28*, 856−862.
(600) Zeitoun, R. I.; Garst, A. D.; Degen, G. D.; Pines, G.; Mansell, T. J.; Glebes, T. Y.; Boyle, N. R.; Gill, R. T. Multiplexed Tracking of Combinatorial Genomic Mutations in Engineered Cell Populations. *Nat. Biotechnol.* **2015**, *33*, 631−637.
(601) Trudeau, D. L.; Tawfik, D. S. Protein Engineers Turned Evolutionists-the Quest for the Optimal Starting Point. *Curr. Opin. Biotechnol.* **2019**, *60*, 46−52.
(602) Peisajovich, S. G.; Tawfik, D. S. Protein Engineers Turned Evolutionists. *Nat. Methods* **2007**, *4*, 991−994.
(603) Bershtein, S.; Segal, M.; Bekerman, R.; Tokuriki, N.; Tawfik, D. S. Robustness−Epistasis Link Shapes the Fitness Landscape of a Randomly Drifting Protein. *Nature* **2006**, *444*, 929−932.
(604) Wagner, A. *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*; OUP Oxford, 2011.
(605) Dellus-Gur, E.; Toth-Petroczy, A.; Elias, M.; Tawfik, D. S. What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-Offs. *J. Mol. Biol.* **2013**, *425*, 2609−2621.
(606) Smith, B. A.; Mularz, A. E.; Hecht, M. H. Divergent Evolution of a Bifunctional De Novo Protein. *Protein Sci.* **2015**, *24*, 246−252.
(607) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D. Evolution of a Highly Active and Enantiospecific Metalloenzyme from Short Peptides. *Science* **2018**, *362*, 1285−1288.
(608) Travisano, M.; Mongold, J. A.; Bennett, A. F.; Lenski, R. E. Experimental Tests of the Roles of Adaptation, Chance, and History in Evolution. *Science* **1995**, *267*, 87−90.
(609) Paaby, A. B.; Rockman, M. V. Cryptic Genetic Variation: Evolution's Hidden Substrate. *Nat. Rev. Genet.* **2014**, *15*, 247−258.
(610) Frankel, N.; Davis, G. K.; Vargas, D.; Wang, S.; Payre, F.; Stern, D. L. Phenotypic Robustness Conferred by Apparently Redundant Transcriptional Enhancers. *Nature* **2010**, *466*, 490−493.
(611) Burga, A.; Casanueva, M. O.; Lehner, B. Predicting Mutation Outcome from Early Stochastic Variation in Genetic Interaction Partners. *Nature* **2011**, *480*, 250−253.
(612) Bergman, A.; Siegal, M. L. Evolutionary Capacitance as a General Feature of Complex Gene Networks. *Nature* **2003**, *424*, 549−552.
(613) Hayden, E. J.; Ferrada, E.; Wagner, A. Cryptic Genetic Variation Promotes Rapid Evolutionary Adaptation in an RNA Enzyme. *Nature* **2011**, *474*, 92−95.
(614) Zheng, J.; Payne, J. L.; Wagner, A. Cryptic Genetic Variation Accelerates Evolution by Opening Access to Diverse Adaptive Peaks. *Science* **2019**, *365*, 347−353.
(615) Bloom, J. D.; Romero, P. A.; Lu, Z.; Arnold, F. H. Neutral Genetic Drift Can Alter Promiscuous Protein Functions, Potentially Aiding Functional Evolution. *Biol. Direct* **2007**, *2*, 17.
(616) Gupta, R. D.; Tawfik, D. S. Directed Enzyme Evolution via Small and Effective Neutral Drift Libraries. *Nat. Methods* **2008**, *5*, 939−942.
(617) Skinner, M. M.; Terwilliger, T. C. Potential Use of Additivity of Mutational Effects in Simplifying Protein Engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 10753.
(618) Mildvan, A. S. Inverse Thinking About Double Mutants of Enzymes. *Biochemistry* **2004**, *43*, 14517−14520.
(619) Aita, T.; Uchiyama, H.; Inaoka, T.; Nakajima, M.; Kokubo, T.; Husimi, Y. Analysis of a Local Fitness Landscape with a Model of the Rough Mt. Fuji-Type Landscape: Application to Prolyl Endopeptidase and Thermolysin. *Biopolymers* **2000**, *54*, 64−79.
(620) Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 138−174.
(621) Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. *J. Am. Chem. Soc.* **2010**, *132*, 9144−9152.
(622) Reetz, M. T.; Sanchis, J. Constructing and Analyzing the Fitness Landscape of an Experimental Evolutionary Process. *ChemBioChem* **2008**, *9*, 2260−2267.
(623) Huang, P. S.; Boyken, S. E.; Baker, D. The Coming of Age of De Novo Protein Design. *Nature* **2016**, *537*, 320−327.
(624) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *NeurIPS* **2017**, 5998−6008.
(625) Bryant, D. H.; Bashir, A.; Sinai, S.; Jain, N. K.; Ogden, P. J.; Riley, P. F.; Church, G. M.; Colwell, L. J.; Kelsic, E. D. Deep Diversification of an AAV Capsid Protein by Machine Learning. *Nat. Biotechnol.* **2021**, *39*, 691.
(626) Gurard-Levin, Z. A.; Scholle, M. D.; Eisenberg, A. H.; Mrksich, M. High-Throughput Screening of Small Molecule Libraries Using Samdi Mass Spectrometry. *ACS Comb. Sci.* **2011**, *13*, 347−350.
(627) Peng, T.; Nagy, G.; Trinidad, J. C.; Jackson, J. M.; Pohl, N. L. B. A High-Throughput Mass-Spectrometry-Based Assay for Identifying the Biochemical Functions of Putative Glycosidases. *ChemBioChem* **2017**, *18*, 2306−2311.
(628) Chang, S.-H.; Han, J.-L.; Tseng, S. Y.; Lee, H.-Y.; Lin, C.-W.; Lin, Y.-C.; Jeng, W.-Y.; Wang, A. H. J.; Wu, C.-Y.; Wong, C.-H. Glycan Array on Aluminum Oxide-Coated Glass Slides through Phosphonate Chemistry. *J. Am. Chem. Soc.* **2010**, *132*, 13371−13380.
(629) Chin, J.; Wood, E.; Peters, G. S.; Drexler, D. M. Acoustic Sample Deposition MALDI-MS (ASD-MALDI-MS): A Novel Process Flow for Quality Control Screening of Compound Libraries. *J. Lab. Autom.* **2016**, *21*, 204−207.
(630) Greving, M.; Cheng, X. L.; Reindl, W.; Bowen, B.; Deng, K.; Louie, K.; Nyman, M.; Cohen, J.; Singh, A.; Simmons, B.; et al. Acoustic Deposition with NIMS as a High-Throughput Enzyme Activity Assay. *Anal. Bioanal. Chem.* **2012**, *403*, 707−711.
(631) Morand, K. L.; Burt, T. M.; Regg, B. T.; Chester, T. L. Techniques for Increasing the Throughput of Flow Injection Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 247−252.
(632) Lowe, D. M.; Gee, M.; Haslam, C.; Leavens, B.; Christodoulou, E.; Hissey, P.; Hardwicke, P.; Argyrou, A.; Webster, S. P.; Mole, D. J.; et al. Lead Discovery for Human Kynurenine 3-Monooxygenase by High-Throughput Rapidfire Mass Spectrometry. *J. Biomol. Screening* **2014**, *19*, 508−515.
(633) Rohman, M.; Wingfield, J. In *High Throughput Screening: Methods and Protocols*; Janzen, W. P., Ed.; Springer: New York, NY, 2016.
(634) Benkestock, K.; Van Pelt, C. K.; Åkerud, T.; Sterling, A.; Edlund, P.-O.; Roeraade, J. Automated Nano-Electrospray Mass Spectrometry for Protein-Ligand Screening by Noncovalent Interaction Applied to Human H-FABP and A-FABP. *J. Biomol. Screening* **2003**, *8*, 247−256.
(635) Schwudke, D.; Hannich, J. T.; Surendranath, V.; Grimard, V.; Moehring, T.; Burton, L.; Kurzchalia, T.; Shevchenko, A. Top-Down Lipidomic Screens by Multivariate Analysis of High-Resolution Survey Mass Spectra. *Anal. Chem.* **2007**, *79*, 4083−4093.
(636) Schebb, N. H.; Vielhaber, T.; Jousset, A.; Karst, U. Development of a Liquid Chromatography-Based Screening Methodology for Proteolytic Enzyme Activity. *J. Chromatogr. A* **2009**, *1216*, 4407−4415.
(637) Beneito-Cambra, M.; Gareil, P.; Badet, B.; Badet-Denisot, M.-A.; Delaunay, N. First Investigations for the Characterization of Glucosamine-6-Phosphate Synthase by Capillary Electrophoresis. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2018**, *1072*, 130−135.
(638) Pei, J. A.; Li, Q. A.; Kennedy, R. T. Rapid and Label-Free Screening of Enzyme Inhibitors Using Segmented Flow Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1107−1113.
(639) Su, Y.; Zhu, Y.; Fang, Q. A Multifunctional Microfluidic Droplet-Array Chip for Analysis by Electrospray Ionization Mass Spectrometry. *Lab Chip* **2013**, *13*, 1876−1882.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444

(640) Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. Tandem Mass Spectrometry for Peptide and Protein Sequence Analysis. *BioTechniques* **2005**, *38*, 519−523.
(641) Musharraf, S. G.; Bhatti, M. S.; Choudhary, M. I.; Rahman, A. U. Screening of Inhibitors of Angiotensin-Converting Enzyme (ACE) Employing High Performance Liquid Chromatography-Electrospray Ionization Triple Quadrupole Mass Spectrometry (HPLC-ESI-QqQ-MS). *Eur. J. Pharm. Sci.* **2017**, *101*, 182−188.
(642) Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N. Fourier Transform Ion Cyclotron Resonance (FT ICR) Mass Spectrometry: Theory and Simulations. *Mass Spectrom. Rev.* **2016**, *35*, 219−258.
(643) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61−80.
(644) Ràfols, P.; Vilalta, D.; Brezmes, J.; Cañellas, N.; del Castillo, E.; Yanes, O.; Ramírez, N.; Correig, X. Signal Preprocessing, Multivariate Analysis and Software Tools for MA(LDI)-TOF Mass Spectrometry Imaging for Biological Applications. *Mass Spectrom. Rev.* **2018**, *37*, 281−306.
(645) Comi, T. J.; Neumann, E. K.; Do, T. D.; Sweedler, J. V. Microms: A Python Platform for Image-Guided Mass Spectrometry Profiling. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1919−1928.
(646) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; et al. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016**, *13*, 741−748.
(647) HamediRad, M.; Chao, R.; Weisberg, S.; Lian, J.; Sinha, S.; Zhao, H. Towards a Fully Automated Algorithm Driven Platform for Biosystems Design. *Nat. Commun.* **2019**, *10*, 5150.
(648) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369−387.
(649) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188−1190.
(650) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. PoPMuSiC 2.1: A Web Server for the Estimation of Protein Stability Changes Upon Mutation and Sequence Optimality. *BMC Bioinf.* **2011**, *12*, 151.
(651) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated Analysis of Interatomic Contacts in Proteins. *Bioinformatics* **1999**, *15*, 327−332.
(652) Laurie, A. T.; Jackson, R. M. Q-SiteFinder: An Energy-Based Method for the Prediction of Protein-Ligand Binding Sites. *Bioinformatics* **2005**, *21*, 1908−1916.
(653) Parthiban, V.; Gromiha, M. M.; Schomburg, D. CUPSAT: Prediction of Protein Stability Upon Point Mutations. *Nucleic Acids Res.* **2006**, *34*, W239−242.
(654) Hernandez, M.; Ghersi, D.; Sanchez, R. SITEHOUND-web: A Server for Ligand Binding Site Identification in Protein Structures. *Nucleic Acids Res.* **2009**, *37*, W413−W416.
(655) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids. *Nucleic Acids Res.* **2010**, *38*, W529−W533.
(656) Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand−Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778−2786.
(657) Tian, J.; Wu, N.; Chu, X.; Fan, Y. Predicting Changes in Protein Thermostability Brought About by Single- or Multi-Site Mutations. *BMC Bioinf.* **2010**, *11*, 370.
(658) Giollo, M.; Martin, A. J. M.; Walsh, I.; Ferrari, C.; Tosatto, S. C. E. NeEMO: A Method Using Residue Interaction Networks to Improve Prediction of Protein Stability Upon Mutation. *BMC Genomics* **2014**, *15*, S7.
(659) Zaugg, J.; Gumulya, Y.; Malde, A. K.; Bodén, M. Learning Epistatic Interactions from Sequence-Activity Data to Predict Enantioselectivity. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1085−1096.
(660) Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity. *arXiv:1703.10603* **2017**.
(661) Mazzaferro, C. Predicting Protein Binding Affinity with Word Embeddings and Recurrent Neural Networks. *bioRxiv* **2017**, 128223.
(662) Sønderby, S. K.; Winther, O. Protein Secondary Structure Prediction with Long Short Term Memory Networks. *arXiv:1412.7828* **2014**.
(663) Zhou, J.; Troyanskaya, O. Deep Supervised and Convolutional Generative Stochastic network for Protein Secondary Structure Prediction. *International conference on machine learning* **2014**, 745−753.
(664) Hwang, B. Y.; Oh, J. M.; Kim, J.; Kim, B. G. Pro-Antibiotic Substrates for the Identification of Enantioselective Hydrolases. *Biotechnol. Lett.* **2006**, *28*, 1181−1185.
(665) Chen, Z.; Zhao, H. Rapid Creation of a Novel Protein Function by in Vitro Coevolution. *J. Mol. Biol.* **2005**, *348*, 1273−1282.
(666) Tyagi, V.; Bonn, R. B.; Fasan, R. Intermolecular Carbene S−H Insertion Catalysed by Engineered Myoglobin-Based Catalysts. *Chem. Sci.* **2015**, *6*, 2488−2494.
(667) Jones, D. D.; Arpino, J. A. J.; Baldwin, A. J.; Edmundson, M. C. In *Directed Evolution Library Creation*; Gillam, E. M. J.; Copp, J. N.; Ackerley, D., Eds.; Springer New York: New York, NY, 2014; Vol. *1179*.
(668) Lian, J.; Schultz, C.; Cao, M.; HamediRad, M.; Zhao, H. Multi-functional Genome-wide CRISPR System for High Throughput Genotype-phenotype Mapping. *Nat. Commun.* **2019**, *10*, 5794.

https://doi.org/10.1021/acs.chemrev.1c00260
Chem. Rev. 2021, 121, 12384−12444