

Received: 15 February 2021 | Revised: 28 May 2021 | Accepted: 2 June 2021

DOI: 10.1002/bies.202100052

METHODS, MODELS & TECHNIQUES

Prospects & Overviews

Methods for enzyme library creation: Which one will you choose?

A guide for novices and experts to introduce genetic diversity

Lorea Alejandre^{1,2}  | Joelle N. Pelletier^{1,2,3}  | Daniela Quaglia^{3,4} 

¹ Département de biochimie and Center for Green Chemistry and Catalysis (CGCC), Université de Montréal, Montréal, Quebec, Canada

² PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, Quebec, Canada

³ Département de chimie, Université de Montréal, Montréal, Quebec, Canada

⁴ School of Chemistry, University of Nottingham, Nottingham, UK

Correspondence

Email: daniela.quaglia1@nottingham.ac.uk, daniela.quaglia@umontreal.ca

Funding information

Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: PVX20965-(RGP)

Abstract

Enzyme engineering allows to explore sequence diversity in search for new properties. The scientific literature is populated with methods to create enzyme libraries for engineering purposes, however, choosing a suitable method for the creation of mutant libraries can be daunting, in particular for the novices. Here, we address both novices and experts: how can one enter the arena of enzyme library design and what guidelines can advanced users apply to select strategies best suited to their purpose? Section I is dedicated to the novices and presents an overview of established and standard methods for library creation, as well as available commercial solutions. The expert will discover an up-to-date tool to freshen up their repertoire (Section I) and learn of the newest methods that are likely to become a mainstay (Section II). We focus primarily on in vitro methods, presenting the advantages of each method. Our ultimate aim is to offer a selection of methods/strategies that we believe to be most useful to the enzyme engineer, whether a first-timer or a seasoned user.

KEYWORDS

enzyme engineering, genetic diversity, library creation methods, sequence space exploration, smart library design

INTRODUCTION

The number of combinations of mutations that can theoretically be made in a protein of 100 amino acids works out to more than 10^{130} possible combinations. If we consider that the largest screened libraries contain approximately 10^{15} members,^[1] it is clear that only an infinitesimal part of the vast theoretical sequence space is accessible through experiments and will remain so even as automation increases screening capacity. For this reason, researchers have focused on improving library quality because the “smarter” the library, the greater the odds of identifying the improved variants you are looking for.^[2–6] From the perspective of screening capacity, we consider “small” a library of less than 10^3 members, “medium-sized” one con-

taining between 10^3 and 10^5 , and “large” when the library contains more than 10^5 members. The ability to generate a library that is diverse, while maintaining theoretical library size as small as possible so that it can be thoroughly screened with currently available methods, increases the chance of identifying functional variants.^[7]

Mutagenesis is used to introduce diversity and navigate the sequence space by modifying the wild-type DNA through substitutions, insertions and deletions. Different exploration strategies are available: (i) random mutagenesis; (ii) recombination/gene shuffling and (iii) targeted approaches.^[5,8–15]

Methods based on random mutagenesis generally make use of an error-prone polymerase in a PCR reaction (epPCR) to introduce mutations at random positions within the coding sequence. This is

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *BioEssays* published by Wiley Periodicals LLC

most useful when the structure or the functional regions of the protein we seek to alter are unknown.^[8,12] However, random mutagenesis requires a powerful high-throughput screening method to isolate improved variants because beneficial mutations occur two to three orders of magnitude less frequently than do mutations that are deleterious to function.^[9,16] Furthermore, randomized libraries severely limit the introduction of consecutive mutations in the context of a single round; even with the most modern of approaches, a single round of random mutagenesis seldom suffices to provide significant improvement.^[11,17,18]

Recombination methodologies involve fragmenting mutated copies of the gene of interest and reassembling them to give rise to new combinations of mutations^[8]. Arguably the most popular application of rational recombination is SCHEMA, which involves the guided recombination of fragments of homologous genes from the same family, giving rise to chimeric proteins.^[19] SCHEMA is particularly of interest because it is characterized by somewhat smaller theoretical library sizes ($\sim 10^3$ – 10^4 variants) than other recombination methods^[20–23] while generating diverse libraries with a higher proportion of functional variants than randomization methods.^[8,24]

While these options are powerful and widely adopted, our review emphasizes focused approaches. They involve targeting specific residues or regions for mutation to one or many possibilities, often in a combinatorial manner, or they might imply the use of insertions and deletions in the gene sequence.^[3,18,25–28]

Targeted mutagenesis is limited by the requirement for structural or biophysical knowledge of the enzyme. To address this limitation, it is common to perform exploratory rounds of randomization that inform subsequent rounds of targeted mutagenesis: Frances Arnold once stated that bringing together rational design and directed evolution is the way forward to the generation of quality libraries containing a high number of sequences with the desired properties.^[4,27,29,30] A targeted strategy might include randomization of certain parts of a gene, but not the whole sequence.^[10,12,17,18,31]

Targeted mutagenesis can produce smaller theoretical libraries (~ 10 – 10^4) that may hold a high proportion of functional variants, which in turn translates in a lower screening effort.^[11,17] This has shown to be especially useful in improving specific properties of an enzyme: activity, selectivity and thermal resistance.^[8]

Although of limited scope for sequence diversification, the simplicity of targeted approaches has its benefits, including limiting the number of variants to be screened to find an improved enzyme.

The most basic approach for the generation of targeted mutations is site-directed mutagenesis (SDM), introduced by Nobel laureate Frances Arnold to create a point mutation^[32]. Its derivatives include site-saturation mutagenesis (SSM), discussed in Section I, for introduction of degenerate codons at a specific position^[9,33–35], and combinatorial saturation mutagenesis (CSM, Sections I and II) for mutating more than one site at a time.

Single site saturation libraries: Saturation libraries offer high quality and small size while giving access to greater sequence diversity than do point mutations. However, building saturation libraries one position at a time severely limits exploration of sequence diversity; in particular, it curtails the discovery of positive epistatic effects, where

specific combinations of mutations (whether proximal or distal) give a greater improvement than expected from combining effects of the point mutants.^[36,37] These complex interactions may result from direct interactions between residues or complex interactions involving conformational dynamics or stability.^[37] Epistasis is challenging to foresee, limiting predictability in the exploration of the sequence space.^[38,39]

Combinatorial saturation libraries: To explore epistasis in saturation libraries, combinatorial approaches can be useful. In this case, more than one site is mutated at once and the combination of possible mutations are assessed. Examples of strategies of this kind are ISM (Iterative Saturation Mutagenesis) and its derivatives.

In ISM, sites of interests are chosen (as many as necessary) that generally contain 1 to 3 amino acids (or more). These are then randomized individually using saturation mutagenesis. The best hits selected from the first randomization are subjected to further iterations of randomization with saturation mutagenesis at the other positions.^[5,9,28,39,40] ISM allows for a *fast convergence of beneficial mutations*, mainly because the choice of positions to mutate is informed by prior structural or biochemical knowledge.^[41] The positions to mutate are chosen according to the final goal: B-FIT for a gain in thermostability^[41]; CASTing for improvement of catalytic properties.^[28,35]

A novel iterative methodology (FRISM), inspired by CAST/ISM, includes a rational prediction aspect that restricts the number of mutants to be screened. FRISM has been proven useful for engineering stereoselectivity in enzyme variants.^[5,42,43]

Below, we review a broad selection of methods and strategies to produce the targeted library that best suits your needs. Section I presents an overview of established approaches for standard library creation and popular commercial solutions to speed up the creation of targeted libraries. For those wanting to further explore protein sequence space in a bid to modify enzyme activity more deeply, Section II presents recently reported, innovative approaches to generate more complex, mainly targeted libraries.

SECTION I: ESTABLISHED METHODS FOR STANDARD SITE-DIRECTED AND SATURATION MUTAGENESIS FOR TARGETED LIBRARY CREATION

Site directed mutagenesis (SDM) allows an amino acid in the target protein to be substituted using mutagenic primers. This is easily achieved using commercially available kits such as QuikChange by Agilent (discussed in Section I-B 7). A direct derivative from SDM is site-saturation mutagenesis, commonly referred to as SSM, for mutation at a single site, or CSM (Combinatorial Saturation Mutagenesis) for more than one randomization site.^[8,9,33,35,36] In SSM and CSM, degenerate mutagenic primers introduce the nineteen other amino acids (or a defined subset of them) at each site.

The main advantage of saturation mutagenesis is that it allows for non-conservative codon substitutions that are unlikely to arise by random mutagenesis, giving access to *non-natural evolution pathways*.^[44]

Direct improvement of saturation mutagenesis entails the use of primers that reduce codon degeneracy to tune it to the needs of the

user.^[36] For instance, one can restrict codon redundancy or the presence of stop codons.^[8,33,45] The twenty possible naturally occurring amino acids are encoded by NNN degeneracy that codes for 64 possible codons (including 3 stop codons), where N is either A, T, G, or C. The NNK degeneracy (K = T or G, also refer to the IUPAC nomenclature for standard letters of degeneracy^[46]) limits the number of codons to 32, while encoding all possible amino acids and only 1 stop codon. The library size is reduced which facilitates screening efforts. It is important to note that when NNK is used in a combinatorial manner, the redundancy and library size increase considerably (i.e., for 1 position mutated with NNK, ~34% of hits are redundant, and for 4 positions, library size reaches 10^6 with a redundancy of ~81%), and includes an important fraction of unwanted stop codons.

If the user is instead willing to sacrifice some of the possible amino acids,^[47] NNT and NNG code for 16 possibilities, while NDT (D = A, G, T) provides 12 codons, for an even smaller library.^[36] More recently, Tang et al.^[45] devised a method called SILM (Small Intelligent Library Method)^[8] as an alternative to NNK that includes all possible amino acids but reduces degeneracy. In SILM, four primers (with NDT, VMA, ATG and TGG, where V = A, C, or G and M = A or C at the site to be mutated)^[45,48] are mixed at a specified ratio. Another method, called 22c-trick, reduces degeneracy by combining three primers (NDT, VHG, and TGG, where H = A, C, T) that contain 22 codons that code for 20 amino acids.^[49] For an ever-tighter control over codon degeneracy, several tools are available to the users, such as: the DC-ANALYZER,^[8,45] MDC-Analyzer,^[50] SwiftLib,^[51] or DeCoDe.^[52]

Many methods have been developed to introduce mutations in the targeted gene at specific positions. Here we describe a selection of the most popular methods employed in the field of enzyme engineering, in order of increasing complexity. We also mention methods that are less commonly used but that could provide advantages for users having more expertise in molecular biology or having specific requirements. Most of the methods described in this section can be adapted to perform site-directed mutagenesis (SDM), site-saturation mutagenesis (SSM) or combinatorial site-saturation mutagenesis (CSM). It is important to remember that the efficiency of mutagenesis is often dependent on the availability of easy-to-use cloning and assembly methods, the most popular of which we describe in Section I-B.

A. ESTABLISHED METHODS FOR SDM AND SSM

1. Whole plasmid site-directed mutagenesis for simplicity and speed

The whole plasmid approach to site-directed mutagenesis is the most popular method for substituting one or a few consecutive amino acids in a target enzyme (Table 1). First commercialized by Stratagene in 1996 as the QuikChange kit (see Section I-B 7), the standard protocol relies on the design of overlapping primers containing the desired mutagenic codon(s) (Figure 1).^[53] These are used to amplify the whole plasmid with a high-fidelity DNA polymerase. The resulting PCR product is a mutated and nicked double-stranded DNA. Treatment with DpnI enzyme eliminates non-mutated methylated wild-type

DNA template; transformation into a bacterial host allows repair of the nicked, mutated strand with the recombination machinery of the host. Although straightforward, this results in poor transformation efficiency due to nicked DNA; furthermore wild-type contamination resulting from incomplete DpnI digestion has a higher transformation efficiency, which can increase the cost of identifying mutated variants.

These limitations make the standard method unsuited to the generation of larger libraries ($> 10^4$ transformants, as also reported in the QuikChange Site Directed Mutagenesis Manuals). Other drawbacks include annealing of the complementary primers to each other instead of to the target plasmid (primer-dimers) and poor PCR amplification. Nonetheless, the whole-plasmid approach remains one of the fastest and simplest ways to introduce mutations, making it attractive to novice as well as experienced users. To enhance its benefits, numerous improvements have been made concerning primer design,^[59-61] incorporating DNA insertions or deletions,^[62] enhancing transformation efficiency and including mutations at multiple sites (multi-site directed mutagenesis).^[62,63] A recent, noteworthy improvement for difficult targets (large plasmid or difficult to deal with due to the presence of stable secondary structures) has also been proposed by Li et al.^[64]

2. Whole-plasmid site-directed mutagenesis with template/product modification for improved mutational efficiency

Whole-plasmid site-directed mutagenesis using template/product modification (Figure 1) are not popular amongst novices, as the protocols are lengthier than with other methods. However, they have proven useful to reduce wild-type template contamination, offering a mutational efficiency of 50–100%.^[55,56]

The Kunkel method and its derivatives^[65] require the generation of whole-plasmid single-strand template DNA containing dUTP instead of dTTP. This often implies using a phage to infect *E. coli* cells defective in dUTPase (*dut*⁻) and uracil N-glycosylase (*ung*⁻). The lack of these enzymatic activities allows the cells to propagate the plasmid DNA containing the target gene with uracil residues. Synthesis of a complementary strand using a mutated primer restores dTTP. Upon transformation into bacteria, the original dUTP-containing strands are cleaved, eliminating the wild-type template.

Although the basic method does not involve PCR, a remarkably improved protocol relies on the use of a high-fidelity polymerase that prevents uracil stalling, Taq ligase, in vitro uracil DNA glycosylase and exonuclease III. The exonuclease cleaves the uracil-containing template and non-desired DNA products. This results in the transformation of double-stranded DNA with higher transformation efficiency.^[55] This protocol gives further accessibility to the Kunkel method by providing a description to generate uracil containing templates without the need to use phages.^[55] Other studies to improve the mutation efficiency of Kunkel-based methods have focused on improving conditions for DNA amplification such as primer design, extension time, and annealing temperature^[66] or the use of phi29 polymerase.^[67]

TABLE 1 Selected mutagenesis methods

| Category | Method | Mutagenesis efficiency ^a | Estimated time[h] ^b | Enzymes required | Sub-cloning required | Recommended protocol |
|--|----------------------|-------------------------------------|--------------------------------|--|----------------------|--|
| Whole-plasmid site-directed mutagenesis | | Less than 100% | ~20 | High-fidelity polymerase DpnI | No | Laible and Boonrod ^[54] |
| Whole plasmid site-directed mutagenesis with template/product modification | Pfunkel ^c | 70–100% | ~20 | T4 Polynucleotide kinase High-fidelity polymerase that reads uracil-containing DNA Taq Ligase Uracil DNA glycosylase Exonuclease III | No | Firnberg and Ostermeier ^[55] |
| | USE ^d | 50–90% | ~40 | T4/T7 DNA polymerase T4 DNA ligase Restriction enzyme | No | Forloni, Liu and Wajapeyee ^[56] |
| Site-directed mutagenesis by overlap extension | SOE | > 98% | ~24 | High-fidelity polymerase DpnI (optional) | Yes | Sambrook and Russell ^[57] |
| Inverse PCR | SLIM | 93% | ~20 | Taq DNA polymerase High-fidelity polymerase DpnI | No | Chiu, March, Lee and Tillett ^[58] |

^aMutagenesis efficiency is as calculated in each of the recommended protocols.

^bEstimated time includes transformation and overnight culture.

^cdut⁻/ung⁻: *E. coli* strain required to generate uracil-containing template.

^dmutS: *E. coli* strain required to avoid repair of the mutated template.

Oligonucleotide-directed mutagenesis by elimination of unique restriction site (USE)^[56,68] also relies on whole-plasmid amplification of DNA. It employs a mutagenic primer and a primer that eliminates a unique restriction site in the mutated plasmid. Subsequent digestion with the corresponding restriction enzyme eliminates wild-type template. This product is transformed into a mutS *E. coli* strain deficient in repair of mismatched bases. The obtained colonies are pooled, the plasmid DNA is obtained and digested with the same restriction enzyme to ensure complete elimination of wild-type template, and transformed in a standard *E. coli* strain. In conjunction with the Kunkel-based methods, USE can serve to increase mutational efficiency.^[69]

3. Site-directed mutagenesis by overlap extension for large plasmids or difficult constructs

First described in 1989 as a PCR-based method to introduce mutations, insertions and deletions,^[70] the main advantages of Site-directed mutagenesis by Overlap Extension (SOE) are high product yield and traceability of product formation. It is mostly used for large plasmids, difficult-to-amplify genes or when whole-plasmid amplification consistently fails due to other reasons (i.e. formation of primer dimers). In SOE, overlapping primers containing the mutation(s) are designed, sim-

ilarly to the commercial QuikChange method (see Section I-B 7). These are used separately in two PCR reactions to introduce the mutation(s), paired with a primer complementary to either the 5' or 3' terminus of the gene (Figure 1). A final PCR reaction combining the two mutated fragments and extending to full length with the terminal primers then recreates the whole, mutated gene with efficiencies of up to 98%.^[70] An improvement of this method allows one to carry it out in a single tube: in this instance, a megaprimer carrying the mutation of interest is generated using one mutagenic primer and a non-mutagenic one that starts at either of the gene's termini. The megaprimer is then used to amplify the whole gene in a second PCR step.^[71]

A variation of SOE for more efficient removal of impurities such as template DNA or excess primers is PAGE-mediated Overlap Extension PCR (POEP).^[72] This modification increases mutational efficiency, especially for multiple-site insertions where efficiency of SOE is near 75% (and reaching 100% for POEP).^[72] Improvements for creation of insertions and deletions > 30 bp have also been described^[73], as well as multi-site directed mutagenesis by creating mutagenic fragments with homologous regions that act as megaprimers.^[74–76]

The main limitation of SOE is the general requirement for subsequent subcloning of the mutated DNA fragment in the desired vector. Homologous recombination in vitro and in vivo can alleviate this limitation and will be discussed in Section I-B 6 and 7.

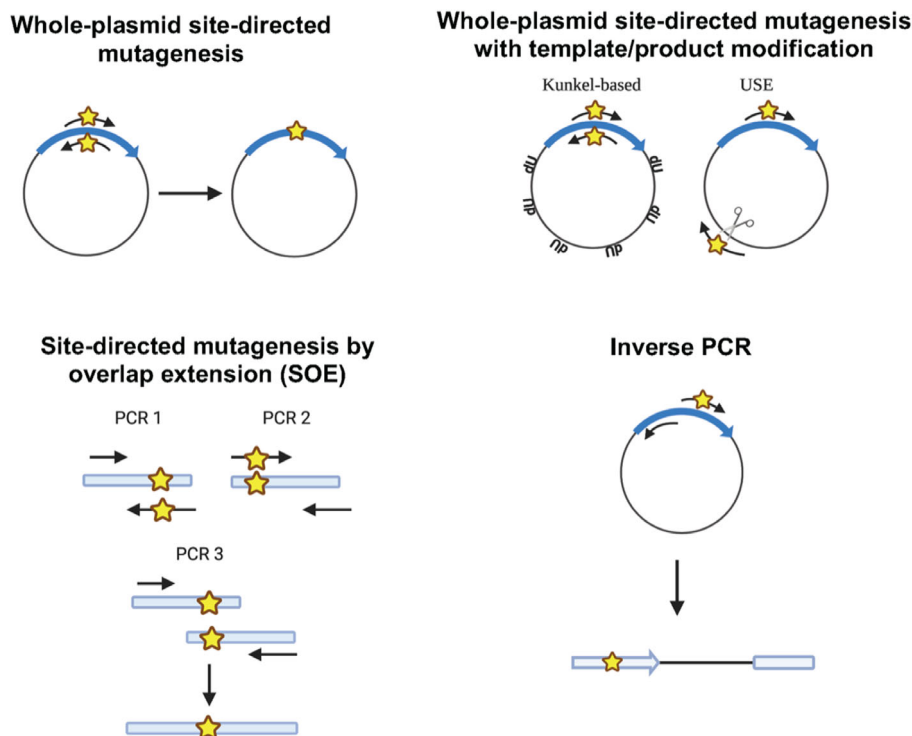


FIGURE 1 Schematic representation of methods described in section I-A. For simplicity schematic representations correspond to double-stranded DNA except for Kunkel-based and USE methodologies where single-stranded DNA is represented. Mutagenic primers and mutated sites are represented with a yellow star, whereas restriction enzymes are represented with scissors. Uracil is indicated by dU on the plasmid. Created in BioRender.com

4. Inverse PCR for saturation mutagenesis and the introduction of deletions and insertions

Developed in 1989,^[77] this approach is particularly advantageous for introducing deletions and insertions, and for saturation mutagenesis.^[78] Inverse PCR relies on the amplification of a targeted sequence using back-to-back primers for outward amplification of regions flanking the targeted sequence (Figure 1). Amplification generates a linear product that is subsequently ligated to regenerate the whole plasmid containing the target gene. Ligation requires a phosphate group at the 5' terminus; this can be achieved with T4 polynucleotide kinase or by use of phosphorylated primers. As with whole-plasmid amplification, digestion of the methylated template DNA with DpnI prior to transformation increases transformation efficiency. An improvement of this method consists in generating complementary overhangs for a ligation-independent approach (SLIM).^[58] A recent improvement (URMAC) is described in Section II-B 2.

B. FLEXIBLE AND EASY-TO-USE ASSEMBLY/CLONING STRATEGIES INSTRUMENTAL TO MUTAGENESIS

The development of straightforward cloning and assembly methodologies in conjunction with site-saturation mutagenesis methods has accelerated the creation of “smarter” methods for library generation.

Sub-cloning of fragments into vectors constitutes a major bottleneck in library generation. This is reflected in the popularity of the QuikChange kit (Section I-B 7) that requires no restriction/ligation step. Ligation-independent methods have provided noteworthy solutions to this problem, accelerating library creation to explore sequence space more efficiently. The major advantage of ligation-independent methods resides in providing increased flexibility in library generation. Mutations introduced in different parts of the genes can subsequently be reassembled to explore interactions between selected mutations or within a defined region in the protein (i.e., a tunnel, the C-terminus, etc.).^[26] A few *ex vivo* and *in vivo* methods will also be mentioned briefly.

1. Restriction-free cloning for versatility

Restriction-Free (RF) cloning can be used to accelerate single or multiple site-directed mutagenesis, insertion and/or deletions.^[79] It consists in amplifying the vector using high-fidelity PCR and an insert (megaprimer) containing regions homologous to the vector in its 3' and 5' termini (Figure 2)^[79,80]. This method is versatile as any vector/insert pair can be used as long as the primer design is carefully performed.^[80] Digestion with DpnI prior to transformation is used to eliminate wild-type template DNA. However, gene replacement of a lethal gene in the original vector can eliminate the need for digestion with DpnI.^[81] The main disadvantages of RF cloning include low

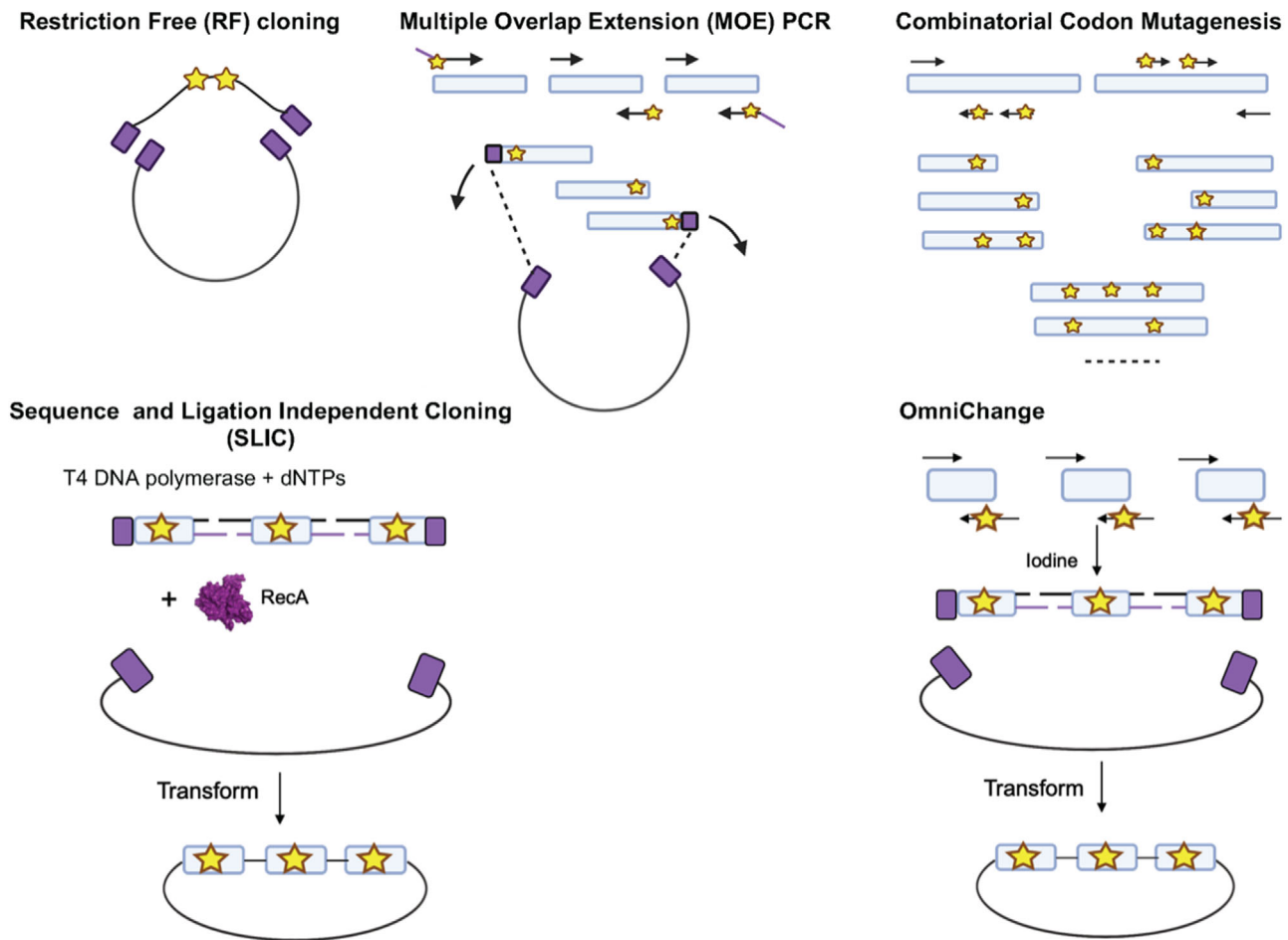


FIGURE 2 Schematic representation of assembly/cloning strategies described in section I-B. Mutations are represented by yellow stars and overlapping segments as purple squares. For simplicity, schematic representations correspond to double-stranded DNA, except in SLIC and OmniChange where nicked double stranded DNA is depicted. Figure inspired by graphical descriptions from references^[74,80,84–86]. Created in BioRender.com

product yield and low efficiency for large insertions. To circumvent these limitations, Exponential Megapriming PCR (EMP) adds forward or reverse primers to the megaprimer reaction to amplify the whole plasmid.^[82] A recent improvement of RF cloning, Modified Restriction-Free cloning (MRF)^[83] allows insertion of fragments of up to 20 kb into the cloning vector.

2. Multiple overlap extension PCR for fragment assembly

Multiple Overlap Extension PCR (MOE-PCR) can be used to assemble up to eight DNA fragments in a single PCR reaction without the need for additional enzymes (T4 DNA ligase, exonuclease, etc.).^[74] In MOE, fragments containing the desired mutations and 50 bp homologous to the contiguous fragment are produced in independent PCR reactions. These are subsequently mixed in a single PCR reaction and used as a template for the assembly of the whole gene or plasmid in a single step (Figure 2). Careful design of the annealing tem-

perature of termini is needed for specificity and efficiency. A previous version of MOE-PCR, called multi-fragment site-directed mutagenic overlap extension PCR, applies the same principle of homologous recombination of fragments termini and amplification: by assembling fragments two by two it can be used to efficiently assemble 13 fragments.^[76]

3. Combinatorial codon mutagenesis based on SOE and megaprimer PCR for tunable mutational frequency

The concept of Combinatorial Codon Mutagenesis (CCM)^[84] is based on SOE-PCR and other recent work.^[87–89] This method allows for the generation of mutant libraries containing between one and seven codon mutations per gene. Two parallel PCR reactions are performed, which contain either a forward primer binding at 5' terminus of the target gene and several mutagenic reverse primers, or a reverse primer binding at 3' terminus and several mutagenic forward primers.

These parallel amplifications generate numerous fragments containing combinations of the mutations introduced in the same PCR reaction (Figure 2). Analogous to SOE, a subsequent PCR containing forward and reverse primers and the generated fragments acting as megaprimers regenerates the whole gene length. The whole gene generated contains different combinations of the mutagenic codons introduced. This method has the advantage of easy “tunability” of mutational frequency as it is a function of fragmentation PCR cycle number (more cycles produce more mutations) and the number of rounds of fragmentation and joining PCRs conducted.

4. Ligation-independent cloning (LIC) and derivatives (SLIC and SLiCE) to ligate inserts into vectors without the need of a ligase

Ligation-Independent Cloning (LIC) allows assembly of insert(s) into vector without ligase as the overhangs created are completely complementary. LIC takes advantage of the exonuclease activity of T4 DNA polymerase to create overhangs in homologous regions between the desired mutated insert(s) and the vector.^[90] This exonuclease activity is favored in the absence of dNTPs and stops when adding specific dNTPs.

Sequence and Ligation-Independent Cloning (SLIC)^[85] (Figure 2), a modification of LIC, allows the recombination of up to 10 fragments. In SLIC, imperfect overhangs are created with T4 DNA polymerase and RecA protein is used to promote homologous recombination *in vitro*. Recombination *ex vivo* using cell lysates of bacteria instead of the RecA protein is possible with a further modification: SLiCE (Seamless Ligation Cloning Extract).^[91] Its application to enzyme engineering^[92] will be described in the *in vivo* and *ex vivo* homology-based recombination methods section I B-6).

The main disadvantage of creating overhangs with T4 DNA polymerase is that ssDNA overhangs created may have strong secondary structures. A recent improvement of this method to recombine short fragments aims at reducing this disadvantage by optimizing incubation temperature and time with T4 DNA polymerase.^[93] A recent variation of SLIC consists in the use of T5 exonuclease instead of T4 DNA polymerase and its transformation in *E. coli* for assembly.^[94]

5. OmniChange for simultaneous mutation in distal parts of the gene

OmniChange can be used to mutagenize up to five codons simultaneously in up to the same number of fragments. As a first step, fragments are created via PCR amplification using forward primers and reverse mutagenic primers, or vice versa. Digestion with DpnI prevents contamination with wild-type template DNA. The mutated fragments are cleaved with an iodine treatment to generate cohesive ends for subsequent reassembly (Figure 2). Nicks are repaired *in vivo* after transformation in *E. coli* BL21-Gold (DE3) IacI^{Q1} (common laboratory strains can be used).

6. *in vivo* and *ex vivo* homology-based recombination

Several methods described above generate nicked products (i.e. whole-plasmid amplification and OmniChange) that are repaired *in vivo* by common laboratory strains of *E. coli*. We deem it useful to briefly enumerate selected examples of easy-to-implement *in vivo* and *ex vivo* (cell lysate) homology-based recombination methods (Table 2). The methods use either *E. coli* or *S. cerevisiae*, the most common organisms for the heterologous expression and directed molecular evolution of prokaryotic and eukaryotic enzymes. We selected two that take advantage of the *recA*-independent recombination pathway of *E. coli* present in common cloning strains such as DH5 α , XL10-Gold or Stbl3: SLiP (*in vivo*)^[91,92] and IVA cloning (*ex vivo*).^[95] REPLACR mutagenesis^[96] instead uses bacteria expressing Red/ET with *recA*-recombineering activity whereas IVOE^[97] uses *S. cerevisiae*. These approaches can be also applied by more experienced users to avoid the cost of commercial kits.

7. Commercial and proprietary methods to facilitate mutagenesis

Several kits are available for mutagenesis and assembly. They can be useful to novices for an easy start or as a standard quick alternative for experienced users. We present a selection of the most widely used kits in Table 3.

Until recently, gene synthesis was not broadly applied to generate libraries of very high quality (namely minimum biases). Nowadays the cost of gene synthesis and its quality have improved making this strategy more accessible. For instance, Twist Bioscience has recently developed a proprietary method for the synthesis of combinatorial libraries that they refer to as *nearing perfection* making commercial gene libraries even more accessible. Their synthesis capability was successfully validated by Li and colleagues, demonstrating how a high-quality combinatorial saturation mutagenesis library can be obtained through high-fidelity solid-phase chemical gene synthesis on silicon chips coupled with gene assembly.^[98] Their method has shown higher mutational efficiency, fewer incomplete sequences and less wild-type contamination than library generation through traditional saturation mutagenesis methods.^[98]

The same laboratory recently presented a second example^[99,100] of how commercially available methods (in this instance from Lab-Genius) can be of aid in the generation of smart libraries. Their method relies on previously developed techniques, in particular DNA assembly and on USER (Uracil-Specific Excision Reagent) cloning and it allows more accurate gene synthesis of *long double stranded polynucleotides*.^[100,101]

While the use of synthetic libraries is no substitute for optimal library design, the steadily decreasing cost of gene synthesis definitely makes it a viable complement to smart library design strategies. Other companies that offer library generation services include ATUM, Codexis, Creative Biogene, Creative Biostructure and ThermoFisher.

TABLE 2 Selected in vivo and ex vivo homology-based recombination methods

| Method | Short Description | Organism |
|---|---|--|
| SLiP - SLiCE-mediated PCR-based site-directed mutagenesis ^[91,92] | Complementary to SOE, the target gene is amplified in two separate PCR reactions, generating two gene fragments containing the desired mutation(s) and short (15–20 bp) overlapping regions to each other and the receiving vector. Assembly of the linearized vector and the mutated fragment is done ex vivo by using SLiCE extracts. | <i>E. coli</i> lysate using the recA-independent recombination pathway (SLiCE extract)(common cloning strains, such as DH5 α , XL10-Gold, or XL1-Blue). |
| IVA cloning – in vivo assembly ^[95] | Fragments containing short (15-20 bp) 5' and 3' overlaps are directly transformed for in vivo assembly. Used to do assembly, mutagenesis, insertions and deletions. | <i>E. coli</i> cells using the recA-independent recombination pathway (common cloning strains, such as DH5 α , XL10-Gold, or XL1-Blue). |
| IVOE - in vivo overlap extension ^[97] | Complementary to SOE, the target gene is amplified in two separate PCR reactions, generating two gene fragments containing the desired mutation(s) and 40–66 bp 5' and 3' overlaps with each other and the linearized receiving vector. These are subsequently transformed for in vivo assembly to regenerate the full gene-length and sub-clone to the vector. | Protease-deficient <i>S. cerevisiae</i> strain BJ 5465, ATCC 208289. |
| REPLACR mutagenesis - Recombineering of Ends of linearised PLAsmids after PCR ^[96] | The vector is linearized by inverse PCR with primers designed to generate short (~17 bp) overlaps at 5' and 3'. The linear PCR product is directly transformed for assembly in vivo. The method can be used to do site-directed mutagenesis, insertions and deletions. | Bacteria expressing Red/ET with recA ⁻ recombineering activity. |

SECTION II: OUR SELECTION OF THE MOST CREATIVE OUT-OF-THE-BOX STRATEGIES FOR LIBRARY CREATION

Below, we review a selection of recent (2016 to 2021), innovative methods that either rethink and improve established methods or propose completely new solutions for “smarter” library generation (refer also to Table 4). The novelty might come in the form of a simpler way to target specific segments of the target for mutagenesis or of a strategy to explore sequence space from a different perspective. We are confident that, if you are dealing with a complex problem of library generation, you will find a solution here or be inspired by one of these methods.

A. ASSEMBLY-BASED STRATEGIES

As mentioned in section I-B, in general, DNA assembly methods can be used as a complementary tool to generate targeted libraries. Some of the methods highlighted in this section belong to this category.

1. Golden Gate-based approaches: treating the enzymes in “parts”

We recently developed a strategy to provide flexibility in blending mutated and non-mutated, user-specified segments of a protein of interest.^[26,102] This approach is based on the Golden-Gate assembly

method, which makes use of Type II restriction enzymes that cut outside of their recognition site, allowing for seamless assembly of fragments.^[110] Among its strengths (see details in Table 4), this method is quick, low-cost and allows for the introduction of highly mutated gene segments (or “parts”) into the wild-type background. It also lends itself to automation and allows for the exploration of epistatic effects.

The main benefit of this approach is to tailor the mutagenic treatment of separate regions within a gene. In our case, a large, putative substrate-binding tunnel was well suited to random mutagenesis, whereas individual residues in other regions were targets for saturation mutagenesis.^[102] The problem was treated by applying the concept of “parts” often used in synthetic biology to assemble gene pathways. The gene was ordered in 3 parts, each designed to include 5' and 3' ends for seamless and easy (ideally one-pot) reconstitution of the whole gene.^[110] Each part was subjected to the appropriate mutagenesis strategy and the mutated and non-mutated parts were reassembled at will to generate a series of gene libraries with various parts – or all parts – mutated. Demonstrated in our work with a gene segmented into 3 parts, we believe the Golden Gate strategy could be applied to at least 25 parts.^[111] Because the user selects the appropriate mutagenesis technology for each part, there is no limit to which options and residues are chosen.

Pullmann et al. followed up with a similar approach to generate multisite saturation mutagenesis libraries, additionally offering a software package to facilitate primer design (<https://msbi.ipb-halle.de/GoldenMutagenesisWeb/>) and a colorimetric control screen for restriction efficiency.^[103]

TABLE 3 Selected commercial mutagenesis and assembly kits

| Goal | Product name | Company | Cost per reaction ^a | Estimated Time ^b [h] | Number of transformants (t) or fragments (f) |
|--|---|---------------------|--------------------------------|---------------------------------|---|
| Site directed mutagenesis | QuikChange lighting site-directed mutagenesis kit | Agilent | \$ | 20 | t: 10 ² –10 ⁴ |
| | QuikChange lighting multi site-directed mutagenesis kit | Agilent | \$\$\$ | 20 | t: 10–10 ³ |
| | Q5 site directed mutagenesis kit | New England Biolabs | \$ | 20 | t: 10 ⁴ |
| | GeneArt site-directed mutagenesis system | ThermoFisher | \$\$ | 22 | n/a ^c |
| Random mutagenesis | Genemorph II random mutagenesis kit | Agilent | \$ | 22 | t: 10 ³ –10 ⁶ ^d |
| | Diversify PCR random mutagenesis kit | Takara | \$ | 22 | t: 10 ³ –10 ⁶ ^d |
| Ligation-independent commercialized homology-based recombination methods | GenBuilder DNA assembly cloning kits | GenScript | \$ | 17 | t: n/a ^c / f: 1 to 12 |
| | NEBuilder HiFi DNA assembly cloning kit | New England Biolabs | \$ | 19 | t: 10 ³ –10 ⁴ /f: (1 to 11) |
| | NEB Gibson assembly cloning kits | New England Biolabs | \$ | 19 | t: 10 ³ –10 ⁴ / f:(1 to 6) |
| | NEB golden gate assembly kits | New England Biolabs | \$ | 19 | t: 10 ³ –10 ⁶ /f: (1 to 24) |
| | In-Fusion HD cloning kit | Takara | \$ | 17 | t: 10 ² –10 ³ /f: (1 to 2) |
| | GeneArt Type IIs assembly kits | ThermoFisher | \$\$ | 20 | t:10 ² –10 ⁵ /f: (1 to 8) |
| | GeneArt™ Gibson assembly EX cloning kit | ThermoFisher | \$ | 20 | t: 10 ² –10 ³ /f: (1 to 15) |
| | GeneArt™ Gibson assembly HiFi cloning kit | ThermoFisher | \$ | 19 | t:10 ³ /f: (1 to 6) |

^aCost per reaction based on the average of the formats available and compared within each category. Prices vary from 5–60 USD per reaction.

^bEstimated time includes transformation and overnight culture.

^cn/a Information not available.

^dDepending on cloning efficiency.

2. Simultaneous introduction of multiple mutations in distal parts of the protein: Darwin Assembly, RECODE

Darwin assembly is a method by Cozens and Pinheiro. It is noteworthy as it represents a relatively simple, fast, and low-cost solution to the introduction of multiple mutations in distal sites of a protein.^[104] Among its strengths (Table 4), it allows for flexibility in the evolutionary pathway that the researcher uncovers: non-contiguous mutations are introduced simultaneously and not sequentially so the evolutionary pathway chosen does not rely on the very first mutation introduced, as per other methods. The method lends itself to automation, for instance by using the Antha software of Synthace.^[112]

Darwin Assembly enables the parallel introduction of more than 10 distal mutations at once and the generation of high-quality libraries (wild-type contamination in the best-case scenario is 0.25%, generally < 0.5%) of > 10⁸ transformants. The authors demonstrated the utility of this method by mutating 19 codons (38 mutations) in a sin-

gle reaction while performing directed evolution on a DNA polymerase. The method was tested for assemblies of up to 2.7 kb.

The three phases involve single stranded DNA generation, isothermal assembly and PCR amplification and cloning. Two versions of the method are shown: one uses biotinylated primers and the second uses a theta primer for the recovery of the mutagenic library and removal of the wild type. The authors applied the method to (i) reassigning a codon in a gene (i.e. they reassigned leucine codons CAT to CTG in the small, 660 bp, gene coding for chloramphenicol acetyl transferase); (ii) generate libraries for amino acid scanning; (iii) generate indel libraries. In this last case, optimization is required as some biases are introduced.

The RECODE^[89] (Rapidly Efficient Combinatorial Oligonucleotides for Directed Evolution) method can similarly be used to generate either single mutations at distal positions, insertions, deletions, and multiple cassettes. This method is similar in nature to the Darwin assembly in that it makes use of several mutagenic primers that are annealed at once on the template DNA albeit different conditions are used. In

RECODE the authors also use *ad-hoc* “anchor” primers similarly to the Darwin assembly. These are designed and standardized to allow facile reassembly of the library into the vector, but they are not biotinylated.

The method is straightforward: the primers are annealed to the template DNA upon denaturation, the nicks between them are then filled and ligated thanks to the action of both enzymes, and double stranded DNA is finally regenerated. The ultimate goal is the achievement of *combinatorial mutations at multiple sites* using one or two-step PCRs. The so-obtained linear DNA can then be cloned into a vector (the authors used Gibson assembly).^[105] A drawback of RECODE compared to Darwin assembly might be higher contamination with wild-type template since the biotin-based purification is not used.

B. OTHER METHODS

1. URMAC for targets of very big dimensions and GC-rich targets

URMAC (UnRestricted Mutagenesis and Cloning)^[106] offers several advantages: it is well suited for large plasmids, having been validated on targets up to 17 kb, and it can be used for deletions, insertions and substitutions. URMAC allows for the introduction of simultaneous mutations and is applicable to GC rich templates, which are often troublesome. These advantages all result from not having to generate a copy of the whole plasmid as in most commonly used methods, which also translates into working with smaller constructs.

Expected products are obtained with a success rate of 95%. URMAC is based on the use of primers and DNA ligation to open and close plasmid templates while simultaneously inserting mutations. The sets of primers used are called “starter primers” and “opener (mutagenic) primers and they are phosphorylated, which allows for their participation in the ligation steps. The main shortcoming of this method, compared for instance to Golden Gate-derived methods (Section II A-1), is the requirement for unique restriction sites. This can become a limitation for plasmids larger than 30 kb. In such an instance, the authors suggest that further recombination methodologies should be considered to transfer back the mutated constructs in the original plasmid.^[106]

2. Nicking mutagenesis for robustness and reliability in the generation of saturation mutagenesis libraries

“Nicking mutagenesis”^[107] is a robust and reliable one-pot method that can be used to generate single or multi-site saturation mutagenesis libraries. It can be used on any plasmid provided that it contains a BbvCI restriction site.

First, a plasmid containing the gene of interest is nicked on one strand using BbvCI and Nb.BbvCI endonuclease, followed by exonuclease III digestion of that strand. A mutagenic strand is then synthesized, to replace the digested strand; this step uses low concentration mutagenic primers and thermal cycling, followed by a reaction with Phusion polymerase. The new, nicked strand is repaired with Taq DNA ligase,

yielding a new, heteroduplex DNA. The non-mutated strand is nicked and digested using a second restriction endonuclease (Nb.BbvCII) and exonuclease III, followed by the synthesis of the complementary strand using a secondary primer and treatment with DpnI to eliminate any wild-type plasmid.

The authors generated libraries of single-site saturation mutagenesis for 71-codon stretches, using a mixture of 71 oligo sets, each carrying NNN degeneracy at one of the 71 codons. The results were analyzed by deep sequencing.

The method was also validated using larger constructs; both single and multi-site protocols were designed.^[107] The report is remarkable for the methods proposed and for its rigor, the method having been corroborated by an external research group to evaluate its robustness and reproducibility. The cost and time for the application of the method was estimated to be, respectively, \$455 and 1 day.

3. CRISPR-Cas9-based methods (ICM): for higher specificity and lower presence of wild-type construct or incorrect sequences

Whereas the CRISPR-Cas9 system is now widely used for editing genomes *in vivo*, there has been little reported of its use *in vitro*. To the best of our knowledge, the paper by She et al.^[108] is one of the first – if not the first – example of applying CRISPR-Cas9 to generate “smart” libraries for enzyme engineering. The ICM method (In vitro CRISPR/Cas9 mediated mutagenesis) was demonstrated for single- or multiple-site directed mutagenesis and for the creation of NNK-based saturation mutagenesis libraries.

The authors directly compared this method with traditional PCR-based saturation mutagenesis approaches, demonstrating its superiority in terms of codons identified and presence of wild-type construct or incorrect sequences.^[108] Further advantages relative to PCR-based methods include flexibility, as it can be applied to almost any scaffold, and a greatly lowered error rate (Table 4).

4. TRIAD and InDel-assembly: for a smoother introduction of insertions and deletions

While Nature often uses DNA insertions and deletions as an evolutionary tool, their use in the lab applied to enzyme evolution is not yet widespread. Methods for the random combinatorial incorporation of InDels (Insertions or Deletions) are available, they are limited with respect to the quality and diversity of the generated library. When these modifications are made instead at targeted sites,^[62,78,58,95,96,106] the exploration of the sequence space is limited.

Because of its uniqueness as an exploratory tool for protein evolution,^[113] even though the insertion of randomized InDels cannot be strictly classified as a focused mutation approach. We believe that it deserves a place in this review

The most commonly encountered problem in the development of methods for InDels is frame-shifting^[6,114–120] which causes a high

TABLE 4 Strengths and limitations of recent methods (section II) to build complex libraries

| Method | Strengths | Limitations |
|--|--|---|
| Golden Gate-based approach 1 2017 ^[102] | (i) Upon receiving the synthetic DNA constructs, the one-pot version allows library generation in less than a day and using commercial reagents that are readily available reagents; (ii) combinatorial; (iii) can be carried out one-pot for simple assemblies; (iii) allows the introduction of highly mutated portions of the gene into a wild-type background; (iv) automatable; (v) allows the exploration of epistatic effects. | (i) Conceptually complex for the beginner; design of the genetic parts can be optimized with company if ordering genes or parts to assemble. |
| Golden Gate-based approach 2 2019 ^[103] | In addition to the points made above (Golden Gate-based approach I), online tool available to facilitate primer design. | Same as above (Golden Gate-based approach 1). |
| Darwin Assembly 2018 ^[104] | (i) From plasmid to DNA library within a day and using commercial reagents that are readily available; (ii) can be carried out one-pot; (iii) can target multiple distal positions in a gene; (iv) reduced wild-type contamination; (v) automatable; (vi) allows the exploration of epistatic effects. | (i) Conceptually complex for the beginner; (ii) in the first version of the method, biotinylated primers need to be used. |
| RECODE 2016 ^[89] | (i) Can be performed in under a day; (ii) can be used either to insert single mutations, insertions, deletions, and multiple cassettes; (iii) involves a simple one step (or two step) PCR; (iv) reduced nonspecific mutations and incorrect ligations thanks to the use of high-fidelity DNA polymerase and ligase. | (i) Uses phosphorylated primers; (ii) needs an extra step via assembly, for instance via the Gibson assembly ^[105] . |
| URMAC 2017 ^[106] | (i) Does not require sub-cloning nor copying the entire plasmid as per other common methods; (ii) can be used for deletions, insertions and substitutions; (iii) well suited for large plasmids (up to 17 kb); (iv) allows for the introduction of simultaneous mutations; (v) applicable in cases of high % GC templates. | Requires the presence of unique restriction sites, which can be hard to find in constructs of > 30 kb. |
| Nicking mutagenesis 2016 ^[107] | (i) Can be performed in less than a day; (ii) it can be applied to any plasmid provided that it contains a BbvCI restriction site; (iii) simple as it can be carried out in one-pot (iv) reliable, robust. | (i) Complexity of designs increases when multiple BbvCI nicking sites are present. |
| CRISPR-Cas9-based method (ICM) 2018 ^[108] | (i) Can be performed in under a day; (ii) flexible, as the cleavage site does not depend on a particular restriction site (the PAM (SpCas9) site occurs on average every 8 to 12 bp); (iii) Cas9 is specific, with no mismatches; (iv) high quality of the generated libraries as it is PCR independent so it avoids the introduction of biases; (vi) efficient for plasmids up to 9 kb. | Requires obtaining the Cas system, or to express the protein. |
| Transposon-based InDels method (TRIAD) 2020 ^[6] | (i) TRIAD taps into solutions that might not be accessible using libraries of point mutants through the efficient random incorporation of InDels (ii) offers the possibility to create libraries focused on a specific region of a protein, (iii) compared to other insertion and deletion methods, it solves the problem of frame shifting, (iv) it does not require the use of highly specialized equipment. | (i) Conceptually complex for the beginner, although simpler than existing methods to introduce InDels; (ii) high-throughput screening may be necessary. |
| InDel-Assembly ^[109] | (i) As for TRIAD, this method taps into solutions that might not be accessible using libraries of point mutants because it allows the random incorporation of InDels in libraries that explore composition and sequence length variation (ii) offers a cost-effective solution for high-quality DNA assembly, (iii) compared to other insertion and deletion methods, it solves the problem of frame shifting, (iv) it does not require the use of highly specialized equipment. | (i) Conceptually complex for the beginner, (ii) it requires the use of paramagnetic streptavidin-coated beads and biotinylated primers. |

frequency of non-functional variants^[6], or the use of highly specialized DNA synthesis equipment.^[121] The use of engineered transposons for sampling InDels has also been explored^[121–124] and methods exist that tackle the frame-shift problems, but are currently limited to generating InDels that are fixed in length and sequence, limited to the generation of deletions, or that are technically difficult.^[6,122–125]

Among the newest methods, two are of relevance in that they: (i) solve the frame-shifting problem, (ii) allow introduction of randomized InDels, (iii) and do not require highly specialized equipment.

The TRIAD method (Transposition-based Random Insertion And Deletion mutagenesis)^[6] overcomes these drawbacks. TRIAD allows one to create libraries characterized by the random insertion of short,

in-frame InDels and can be used to introduce single InDels (of a length of up to three nucleotide triplets, or 9 bp) generating either between-codon or cross-codon mutations without frame-shifting.

The InDel-assembly^[109] is a novel DNA assembly platform that provides a means to introduce InDels of varying length and composition through the use of standardized dsDNA building blocks. InDel-assembly works through cycles of restriction and ligation of the building blocks on paramagnetic beads. In contrast to TRIAD, this method was applied to create a targeted library: for instance, it has been used to engineer protein loops or linkers in a β -lactamase. In addition, the authors also describe an *ad-hoc* strategy to analyze the generated libraries.^[109]

CONCLUSIONS

Identifying a suitable method for the creation of a library for enzyme engineering is not as straightforward as it may seem, in particular for novices. A plethora of established and more recent methods described in the literature make it difficult to determine “*which method suits my needs?*” The overview of the most commonly used methods for the generation of targeted mutant libraries in Section I should facilitate this task and open the door for novice users to enter this area.

While the field of enzyme engineering expands, new approaches addressing *where* to mutate or *how* to mutate genetic sequences are increasingly sought for. The newest and most promising methods for “smart” library creation reviewed in Section II illustrates the creativity with which these challenges are addressed. The “smarter” the method used for generating the library of variants, the more efficient the exploration of theoretical protein sequence space, and the lower the throughput of screening needed to uncover improved variants. Even the most innovative of advances will never enable exploring all of sequence space. Nonetheless, the advent of methods that facilitate generation of libraries with improved quality will make us better at navigating sequence diversity.

Looking to the future necessarily brings computer aided strategies into play, to realize the goal of “*generating small(est) and high(est) quality mutant libraries*”.^[126] The main barrier to using predictive or computational methods is likely the unwillingness of chemists and biologists of “*getting acquainted with computational techniques*”.^[126] Happily, many computational tools for enzyme engineering no longer require being a computational expert.^[127] In the not-too-distant future, machine learning will procure a “*third approach*” together with rational design and directed evolution, to be used in complementary and ever-more-powerful ways.^[128] These approaches may be useful to better target the mutational hot-spots in a protein, hence, together with the selection of a smart method for library construction, they have the potential to reduce even further the number of variants that need to be screened before the improved enzyme can be identified.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were generated or analyzed in this review.

ORCID

Lorea Alejaldre  <https://orcid.org/0000-0003-3086-5446>

Joelle N. Pelletier  <https://orcid.org/0000-0002-2934-6940>

Daniela Quaglia  <https://orcid.org/0000-0003-3756-3782>

REFERENCES

- Baker, M. (2011). Protein engineering: Navigating between chance and reason. *Nature Methods*, 8(8), 623–626. <https://doi.org/10.1038/nmeth.1654>
- Gumulya, Y., Sanchis, J., & Reetz, M. T. (2012). Many pathways in laboratory evolution can lead to improved enzymes: How to escape from local minima. *Chembiochem*, 13(7), 1060–1066. <https://doi.org/10.1002/cbic.201100784>
- Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), 734–743. <https://doi.org/10.1016/j.copbio.2010.08.011>
- Chica, R. A., Doucet, N., & Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology*, 16(4), 378–384. <https://doi.org/10.1016/j.copbio.2005.06.004>
- Qu, G., Li, A., Sun, Z., Acevedo-Rocha, C. G., & Reetz, M. T. (2019). The crucial role of methodology development in directed evolution of selective enzymes. *Angewandte Chemie International Edition*, 59(32), 13204–13231. <https://doi.org/10.1002/anie.201901491>
- Emond, S., Petek, M., Kay, E. J., Heames, B., Devenish, S. R. A., Tokuriki, N., & Hollfelder, F. (2020). Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. *Nature Communications*, 11(1), 3469. <https://doi.org/10.1038/s41467-020-17061-3>
- Bunzel, H. A., Garrabou, X., Pott, M., & Hilvert, D. (2018). Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. *Current Opinion in Structural Biology*, 48, 149–156. <https://doi.org/10.1016/j.sbi.2017.12.010>
- Ruff, A. J., Dennig, A., & Schwaneberg, U. (2013). To get what we aim for—progress in diversity generation methods. *The Febs Journal*, 280(13), 2961–2978. <https://doi.org/10.1111/febs.12325>
- Gillam, E. M., Copp, J. N., & Ackerley, D. (2014). *Directed evolution library creation. Methods and protocols* & Eds.) Vol., Vol. 2, Methods in molecular biology, 1179. New York: Springer.
- Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., & Arnold, F. H. (2005). Evolving strategies for enzyme engineering. *Current Opinion in Structural Biology*, 15(4), 447–452. <https://doi.org/10.1016/j.sbi.2005.06.004>
- Goldsmith, M., & Tawfik, D. S. (2012). Directed enzyme evolution: Beyond the low-hanging fruit. *Current Opinion in Structural Biology*, 22(4), 406–412. <https://doi.org/10.1016/j.sbi.2012.03.010>
- Ali, M., Ishqi, H. M., & Husain, Q. (2020). Enzyme engineering: Reshaping the biocatalytic functions. *Biotechnology and Bioengineering*, 117(6), 1877–1894. <https://doi.org/10.1002/bit.27329>
- Tee, K. L., & Wong, T. S. (2013). Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances*, 31(8), 1707–1721. <https://doi.org/10.1016/j.biotechadv.2013.08.021>
- Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7), 379–394. <https://doi.org/10.1038/nrg3927>

15. Denard, C. A., Ren, H., & Zhao, H. (2015). Improving and repurposing biocatalysts via directed evolution. *Current Opinion in Chemical Biology*, 25, 55–64. <https://doi.org/10.1016/j.cbpa.2014.12.036>
16. Goldsmith, M., & Tawfik, D. S. (2013). Enzyme engineering by targeted libraries. *Methods in Enzymology*, 523, 257–283. <https://doi.org/10.1016/B978-0-12-394292-0.00012-6>
17. Martínez, R., & Schwaneberg, U. (2013). A roadmap to directed enzyme evolution and screening systems for biotechnological applications. *Biological Research*, 46, 395–405. <https://doi.org/10.4067/S0716-97602013000400011>
18. Wong, T. S., Roccatano, D., & Schwaneberg, U. (2007). Steering directed protein evolution: Strategies to manage combinatorial complexity of mutant libraries. *Environmental Microbiology*, 9(11), 2645–2659. <https://doi.org/10.1111/j.1462-2920.2007.01411.x>
19. Voigt, C. A., Martínez, C., Wang, Z. G., Mayo, S. L., & Arnold, F. H. (2002). Protein building blocks preserved by recombination. *Nature Structural Biology*, 9(7), 553–558. <https://doi.org/10.1038/nsb805>
20. Otey, C. R., Landwehr, M., Endelman, J. B., Hiraga, K., Bloom, J. D., & Arnold, F. H. (2006). Structure-guided recombination creates an artificial family of cytochromes P450. *Plos Biology*, 4(5), e112. <https://doi.org/10.1371/journal.pbio.0040112>
21. Meyer, M. M., Silberg, J. J., Voigt, C. A., Endelman, J. B., Mayo, S. L., Wang, Z. G., & Arnold, F. H. (2003). Library analysis of SCHEMA-guided protein recombination. *Protein Science*, 12(8), 1686–1693. <https://doi.org/10.1110/ps.0306603>
22. Ostermeier, M., Shim, J. H., & Benkovic, S. J. (1999). A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology*, 17(12), 1205–1209. <https://doi.org/10.1038/70754>
23. Sieber, V., Martínez, C. A., & Arnold, F. H. (2001). Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnology*, 19(5), 456–460. <https://doi.org/10.1038/88129>
24. Bedbrook, C. N., Rice, A. J., Yang, K. K., Ding, X., Chen, S., LeProust, E. M., Gradinaru, V., & Arnold, F. H. (2017). Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins. *Proceedings of the National Academy of Science of the United States of America*, 114(13), E2624–E2633. <https://doi.org/10.1073/pnas.1700269114>
25. Schmitzer, A. R., Lépine, F., & Pelletier, J. N. (2004). Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: Creating alternative functional configurations. *Protein Engineering, Design & Selection*, 17(11), 809–819. <https://doi.org/10.1093/protein/gzh090>
26. Quaglia, D., Alejaldre, L., Ouadhi, S., Rousseau, O., & Pelletier, J. N. (2019). Holistic engineering of Cal-A lipase chain-length selectivity identifies triglyceride binding hot-spot. *Plos One*, 14(1), e0210100. <https://doi.org/10.1371/journal.pone.0210100>
27. Reetz, M. T., Wilensek, S., Zha, D., & Jaeger, K. E. (2001). Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angewandte Chemie International Edition*, 40(19), 3589–3591. [https://doi.org/10.1002/1521-3773\(20011001\)40:19<3589:AID-ANIE3589>3.0.CO;2-X](https://doi.org/10.1002/1521-3773(20011001)40:19<3589:AID-ANIE3589>3.0.CO;2-X)
28. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D., & Vogel, A. (2005). Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angewandte Chemie International Edition*, 44(27), 4192–4196. <https://doi.org/10.1002/anie.200500767>
29. Williams, G. J., Goff, R. D., Zhang, C., & Thorson, J. S. (2008). Optimizing glycosyltransferase specificity via “hot spot” saturation mutagenesis presents a catalyst for novobiocin glycorandomization. *Chemistry & Biology*, 15(4), 393–401. <https://doi.org/10.1016/j.chembiol.2008.02.017>
30. Cheng, F., Zhu, L., & Schwaneberg, U. (2015). Directed evolution 2.0: Improving and deciphering enzyme properties. *Chemical Communications (Cambridge, England)*, 51(48), 9760–9772. <https://doi.org/10.1039/c5cc01594d>
31. Tracewell, C. A., & Arnold, F. H. (2009). Directed enzyme evolution: Climbing fitness peaks one amino acid at a time. *Current Opinion in Chemical Biology*, 13(1), 3–9. <https://doi.org/10.1016/j.cbpa.2009.01.017>
32. Hutchison, C. A. 3rd, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., & Smith, M. (1978). Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry*, 253(18), 6551–6560.
33. Acevedo-Rocha, C. G., Reetz, M. T., & Nov, Y. (2015). Economical analysis of saturation mutagenesis experiments. *Scientific Reports*, 5(1), 10654. <https://doi.org/10.1038/srep10654>
34. Smith, M. (1994). Nobel lecture. Synthetic DNA and biology. *BioScience Reports*, 14(2), 51–66. <https://doi.org/10.1007/BF01210301>
35. Acevedo-Rocha, C. G., Sun, Z., & Reetz, M. T. (2018). Clarifying the difference between iterative saturation mutagenesis as a rational guide in directed evolution and omnichange as a gene mutagenesis technique. *ChemBiochem*, 19(24), 2542–2544. <https://doi.org/10.1002/cbic.201800372>
36. Siloto, R. M. P., & Weselake, R. J. (2012). Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatalysis and Agricultural Biotechnology*, 1(3), 181–189. <https://doi.org/10.1016/j.bcab.2012.03.010>
37. Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, 25(7), 1204–1218. <https://doi.org/10.1002/pro.2897>
38. Miton, C. M., & Tokuriki, N. (2016). How mutational epistasis impairs predictability in protein evolution and design. *Protein Science*, 25(7), 1260–1272. <https://doi.org/10.1002/pro.2876>
39. Reetz, M. T. (2013). The importance of additive and non-additive mutational effects in protein engineering. *Angewandte Chemie International Edition*, 52(10), 2658–2666. <https://doi.org/10.1002/anie.201207842>
40. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., & Bocola, M. (2010). Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. *Journal of the American Chemical Society*, 132, 9144–9152.
41. Reetz, M. T., & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols*, 2(4), 891–903. <https://doi.org/10.1038/nprot.2007.72>
42. Li, D., Wu, Q., & Reetz, M. T. (2020). Focused rational iterative site-specific mutagenesis (FRISM). *Methods in Enzymology*, 643, 225–242. <https://doi.org/10.1016/bs.mie.2020.04.055>
43. Xu, J., Cen, Y., Singh, W., Fan, J., Wu, L., Lin, X., Zhou, J., Huang, M., Reetz, M. T., & Wu, Q. (2019). Stereodivergent protein engineering of a lipase to access all possible stereoisomers of chiral esters with two stereocenters. *Journal of the American Chemical Society*, 141(19), 7934–7945. <https://doi.org/10.1021/jacs.9b02709>
44. Miyazaki, K., & Arnold, F. H. (1999). Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *Journal of Molecular Evolution*, 49(6), 716–720. <https://doi.org/10.1007/PL00006593>
45. Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M., & Jiang, R. (2012). Construction of “small-intelligent” focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques*, 52(3), 149–158. <https://doi.org/10.2144/000113820>
46. Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendations 1984. *Nucleic Acids Research*, 13(9), 3021–3030. <https://doi.org/10.1093/nar/13.9.3021>
47. Li, A., Qu, G., Sun, Z., & Reetz, M. T. (2019). Statistical analysis of the benefits of focused saturation mutagenesis in directed evolution based on reduced amino acid alphabets. *ACS Catalysis*, 9(9), 7769–7778. <https://doi.org/10.1021/acscatal.9b02548>
48. Gaytan, P., & Roldan-Salgado, A. (2013). Elimination of redundant and stop codons during the chemical synthesis of degenerate

- oligonucleotides. Combinatorial testing on the chromophore region of the red fluorescent protein mKate. *ACS Synthetic Biology*, 2(8), 453–462. <https://doi.org/10.1021/sb3001326>
49. Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z.-G., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2012). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synthetic Biology*, 2(2), 83–92. <https://doi.org/10.1021/sb300037w>
 50. Tang, L., Wang, X., Ru, B., Sun, H., Huang, J., & Gao, H. (2014). MDC-Analyzer: A novel degenerate primer design tool for the construction of intelligent mutagenesis libraries with contiguous sites. *Biotechniques*, 56(6), 301–302, 304, 306–308. <https://doi.org/10.2144/000114177>
 51. Jacobs, T. M., Yumerefendi, H., Kuhlman, B., & Leaver-Fay, A. (2015). SwiftLib: Rapid degenerate-codon-library optimization through dynamic programming. *Nucleic Acids Research*, 43(5), e34. <https://doi.org/10.1093/nar/gku1323>
 52. Shimko, T. C., Fordyce, P. M., & Orenstein, Y. (2020). DeCoDe: Degenerate codon design for complete protein-coding DNA libraries. *Bioinformatics*, 36(11), 3357–3364. <https://doi.org/10.1093/bioinformatics/btaa162>
 53. Papworth, C., Bauer, J. C., & Braman, J. C., & (1996). QuickChange site-directed mutagenesis. *Strategies*, 9(8), 3–4. <https://doi.org/10.1080/08924562.1996.11000299>
 54. Laible, M., & Boonrod, K. (2009). Homemade site directed mutagenesis of whole plasmids. *Journal of Visualized Experiments: JoVE*, 27, e1135. <https://doi.org/10.3791/1135>
 55. Firnberg, E., & Ostermeier, M. (2012). PFunkel: Efficient, expansive, user-defined mutagenesis. *Plos One*, 7(12), e52031. <https://doi.org/10.1371/journal.pone.0052031>
 56. Forloni, M., Liu, A. Y., & Wajapeyee, N. (2019). Oligonucleotide-directed mutagenesis by elimination of a unique restriction site (USE mutagenesis). *Cold Spring Harbor Protocols*, (1), pdb.prot097782. <https://doi.org/10.1101/pdb.prot097782>
 57. Sambrook, J., & Russell, D. W. (2006). Site-specific mutagenesis by overlap extension. *Cold Spring Harbor Protocols*, 1(1), pdb.prot3468. <https://doi.org/10.1101/pdb.prot3468>
 58. Chiu, J., March, P. E., Lee, R., & Tillett, D. (2004). Site-directed, Ligase-Independent Mutagenesis (SLIM): A single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Research*, 32(21), e174. <https://doi.org/10.1093/nar/gnh172>
 59. Zheng, L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 32(14), e115. <https://doi.org/10.1093/nar/gnh110>
 60. Miyazaki, K., & Takenouchi, M. (2002). Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *Biotechniques*, 33(5), 1033–1034, 1036–1038. <https://doi.org/10.2144/02335st03>
 61. Tseng, W. C., Lin, J. W., Wei, T. Y., & Fang, T. Y. (2008). A novel megaprimered and ligase-free, PCR-based, site-directed mutagenesis method. *Analytical Biochemistry*, 375(2), 376–378. <https://doi.org/10.1016/j.ab.2007.12.013>
 62. Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *Bmc Biotechnology [Electronic Resource]*, 8, 91. <https://doi.org/10.1186/1472-6750-8-91>
 63. Hogrefe, H. H., Cline, J., Youngblood, G. L., & Allen, R. M. (2002). Creating randomized amino acid libraries with the QuikChange multi site-directed mutagenesis kit. *Biotechniques*, 33(5), 1158–1160, 1162, 1158–1165. <https://doi.org/10.2144/02335pf01>
 64. Li, A., Acevedo-Rocha, C. G., & Reetz, M. T. (2018). Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Applied Microbiology and Biotechnology*, 102(14), 6095–6103. <https://doi.org/10.1007/s00253-018-9041-2>
 65. Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America* 82, 488–492. <https://doi.org/10.1073/pnas.82.2.488>
 66. Liu, B., Long, S., & Liu, J. (2020). Improving the mutagenesis efficiency of the Kunkel method by codon optimization and annealing temperature adjustment. *Nature Biotechnology*, 56, 46–53. <https://doi.org/10.1016/j.nbt.2019.11.004>
 67. Huovinen, T., Brockmann, E.-C., Akter, S., Perez-Gamarra, S., Ylä-Pelto, J., Liu, Y., & Lamminmäki, U. (2012). Primer extension mutagenesis powered by selective rolling circle amplification. *Plos One*, 7(2), e31817. <https://doi.org/10.1371/journal.pone.0031817>
 68. Deng, W. P., & Nickoloff, J. A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Analytical Biochemistry*, 200(1), 81–88. [https://doi.org/10.1016/0003-2697\(92\)90280-k](https://doi.org/10.1016/0003-2697(92)90280-k)
 69. Markvardsen, P., Lassen, S. F., Borchert, V., & Clausen, I. G. (1995). Uracil-USE, an improved method for site-directed mutagenesis on double-stranded plasmid DNA. *Biotechniques*, 18(3), 370–372.
 70. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1), 51–59. [https://doi.org/10.1016/0378-1119\(89\)90358-2](https://doi.org/10.1016/0378-1119(89)90358-2)
 71. Forloni, M., Liu, A. Y., & Wajapeyee, N. (2019). Megaprimer polymerase chain reaction (PCR)-based mutagenesis. *Cold Spring Harbor Protocols*, (6), pdb.prot097824. <https://doi.org/10.1101/pdb.prot097824>
 72. Peng, R. H., Xiong, A. S., & Yao, Q. H. (2006). A direct and efficient PAGE-mediated overlap extension PCR method for gene multiple-site mutagenesis. *Applied Microbiology and Biotechnology*, 73(1), 234–240. <https://doi.org/10.1007/s00253-006-0583-3>
 73. Lee, J., Shin, M. K., Ryu, D. K., Kim, S., & Ryu, W. S. (2010). Insertion and deletion mutagenesis by overlap extension PCR. *Methods in Molecular Biology*, 634, 137–146. https://doi.org/10.1007/978-1-60761-652-8_10
 74. Kadkhodaei, S., Memari, H. R., Abbasiliasi, S., Rezaei, M. A., Movahedi, A., Shun, T. J., & Ariff, A. B. (2016). Multiple overlap extension PCR (MOE-PCR): An effective technical shortcut to high throughput synthetic biology. *RSC Advances*, 6(71), 66682–66694. <https://doi.org/10.1039/C6RA13172G>
 75. Wei, H., Hu, J., Wang, L., Xu, F., & Wang, S. (2012). Rapid gene splicing and multi-sited mutagenesis by one-step overlap extension polymerase chain reaction. *Analytical Biochemistry*, 429(1), 76–78. <https://doi.org/10.1016/j.ab.2012.06.027>
 76. Waneskog, M., & Bjerling, P. (2014). Multi-fragment site-directed mutagenic overlap extension polymerase chain reaction as a competitive alternative to the enzymatic assembly method. *Analytical Biochemistry*, 444, 32–37. <https://doi.org/10.1016/j.ab.2013.09.021>
 77. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., & Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research*, 17(16), 6545–6551. <https://doi.org/10.1093/nar/17.16.6545>
 78. Jain, P. C., & Varadarajan, R. (2014). A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Analytical Biochemistry*, 449, 90–98. <https://doi.org/10.1016/j.ab.2013.12.002>
 79. Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R., & Peleg, Y. (2010). Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *Journal of Structural Biology*, 172(1), 34–44. <https://doi.org/10.1016/j.jsb.2010.06.016>
 80. van den Ent, F., & Löwe, J. (2006). RF cloning: A restriction-free method for inserting target genes into plasmids. *Journal of Biochemical and Biophysical Methods*, 67(1), 67–74. <https://doi.org/10.1016/j.jbbm.2005.12.008>

81. Lund, B., Leiros, H.-K., & Bjerga, G. E. (2014). A high-throughput, restriction-free cloning and screening strategy based on ccdB-gene replacement. *Microbial Cell Factories*, 13(1), 38. <https://doi.org/10.1186/1475-2859-13-38>
82. Ulrich, A., Andersen, K. R., & Schwartz, T. U. (2012). Exponential megaprimering PCR (EMP) cloning—seamless DNA insertion into any target plasmid without sequence constraints. *Plos One*, 7(12), e53360. <https://doi.org/10.1371/journal.pone.0053360>
83. Zeng, F., Hao, Z., Li, P., Meng, Y., Dong, J., & Lin, Y. (2017). A restriction-free method for gene reconstitution using two single-primer PCRs in parallel to generate compatible cohesive ends. *Bmc Biotechnology [Electronic Resource]*, 17(1), 32. <https://doi.org/10.1186/s12896-017-0346-5>
84. Belsare, K. D., Andorfer, M. C., Cardenas, F. S., Chael, J. R., Park, H. J., & Lewis, J. C. (2017). A simple combinatorial codon mutagenesis method for targeted protein engineering. *ACS Synthetic Biology*, 6(3), 416–420. <https://doi.org/10.1021/acssynbio.6b00297>
85. Li, M. Z., & Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods*, 4(3), 251–256. <https://doi.org/10.1038/nmeth1010>
86. Dennig, A., Shivange, A. V., Marienhagen, J., & Schwaneberg, U. (2011). OmniChange: The sequence independent method for simultaneous site-saturation of five codons. *Plos One*, 6(10), e26222. <https://doi.org/10.1371/journal.pone.0026222>
87. Bloom, J. D. (2014). An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular Biology and Evolution*, 31(8), 1956–1978. <https://doi.org/10.1093/molbev/msu173>
88. Heckman, K. L., & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols*, 2(4), 924–932. <https://doi.org/10.1038/nprot.2007.132>
89. Jin, P., Kang, Z., Zhang, J., Zhang, L., Du, G., & Chen, J. (2016). Combinatorial evolution of enzymes and synthetic pathways using one-step PCR. *ACS Synthetic Biology*, 5(3), 259–268. <https://doi.org/10.1021/acssynbio.5b00240>
90. Aslanidis, C., & de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research*, 18, 6069–6074. <https://doi.org/10.1093/nar/18.20.6069>
91. Motohashi, K. (2015). A simple and efficient seamless DNA cloning method using SLiCE from Escherichia coli laboratory strains and its application to SLiP site-directed mutagenesis. *Bmc Biotechnology [Electronic Resource]*, 15(1), 47. <https://doi.org/10.1186/s12896-015-0162-8>
92. Motohashi, K. (2017). Seamless Ligation Cloning Extract (SLiCE) method using cell lysates from laboratory escherichia coli strains and its application to SLiP site-directed mutagenesis. (In: A. Reeves Ed.), *In vitro mutagenesis*. (Vol. 1498, pp. 349–357). New York, NY: Humana Press.
93. Islam, M. N., Lee, K. W., Yim, H.-S., Lee, S. H., Jung, H. C., Lee, J.-H., & Jeong, J.-Y. (2017). Optimizing T4 DNA polymerase conditions enhances the efficiency of one-step sequence- and ligation-independent cloning. *Biotechniques*, 63(3), 3. <https://doi.org/10.2144/000114588>
94. Xia, Y., Li, K., Li, J., Wang, T., Gu, L., & Xun, L. (2019). T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Research*, 47(3), e15. <https://doi.org/10.1093/nar/gky1169>
95. Garcia-Nafria, J., Watson, J. F., & Greger, I. H. (2016). IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Scientific Reports*, 6, 27459. <https://doi.org/10.1038/srep27459>
96. Trehan, A., Kielbus, M., Czapski, J., Stepulak, A., Huhtaniemi, I., & Rivero-Muller, A. (2016). REPLACR-mutagenesis, a one-step method for site-directed mutagenesis by recombineering. *Scientific Reports*, 6, 19121. <https://doi.org/10.1038/srep19121>
97. Alcalde, M., Zumarraga, M., Polaina, J., Ballesteros, A., & Plou, F. (2006). Combinatorial saturation mutagenesis by in vivo overlap extension for the engineering of fungal laccases. *Combinatorial Chemistry & High Throughput Screening*, 9(10), 719–727. <https://doi.org/10.2174/138620706779026079>
98. Li, A., Acevedo-Rocha, C. G., Sun, Z., Cox, T., Xu, J. L., & Reetz, M. T. (2018). Beating bias in the directed evolution of proteins: Combining high-fidelity on-chip solid-phase gene synthesis with efficient gene assembly for combinatorial library construction. *ChemBiochem*, 19(3), 221–228. <https://doi.org/10.1002/cbic.201700540>
99. Li, A., Acevedo-Rocha, C. G., D'Amore, L., Chen, J., Peng, Y., Garcia-Borras, M., Gao, C., Zhu, J., Rickerby, H., Osuna, S., Zhou, J., & Reetz, M. T. (2020). Regio- and stereoselective steroid hydroxylation at the C7-position by cytochrome P450 monooxygenase mutants. *Angewandte Chemie International Edition*, 59(30), 12499–12505. <https://doi.org/10.1002/anie.202003139>
100. Field, J. E. J., & Rickerby, H. F. (2017). WO 2017/046594.
101. Nour-Eldin, H. H., Geu-Flores, F., & Halkier, B. A. (2010). USER Cloning and USER Fusion: The ideal cloning techniques for small and big laboratories. (In: A. Fett-Neto Ed.), *Plant secondary metabolism engineering*. (pp. 185–200). Totowa, NJ: Humana Press.
102. Quaglia, D., Ebert, M. C., Mugford, P. F., & Pelletier, J. N. (2017). Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. *Plos One*, 12(2), e0171741. <https://doi.org/10.1371/journal.pone.0171741>
103. Pullmann, P., Ulpinnis, C., Marillonnet, S., Gruetzner, R., Neumann, S., & Weissenborn, M. J. (2019). Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Scientific Reports*, 9(1), 10932. <https://doi.org/10.1038/s41598-019-47376-1>
104. Cozens, C., & Pinheiro, V. B. (2018). Darwin Assembly: Fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Research*, 46(8), e51. <https://doi.org/10.1093/nar/gky067>
105. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>
106. Hallak, L. K., Berger, K., Kaspar, R., Kwilas, A. R., Montanaro, F., & Peoples, M. E. (2017). Efficient method for site-directed mutagenesis in large plasmids without subcloning. *Plos One*, 12(6), e0177788. <https://doi.org/10.1371/journal.pone.0177788>
107. Wrenbeck, E. E., Klesmith, J. R., Stapleton, J. A., Adeniran, A., Tyo, K. E., & Whitehead, T. A. (2016). Plasmid-based one-pot saturation mutagenesis. *Nature Methods*, 13(11), 928–930. <https://doi.org/10.1038/nmeth.4029>
108. She, W., Ni, J., Shui, K., Wang, F., He, R., Xue, J., Reetz, M. T., Li, A., & Ma, L. (2018). Rapid and error-free site-directed mutagenesis by a PCR-free in vitro CRISPR/Cas9-mediated mutagenic system. *ACS Synthetic Biology*, 7(9), 2236–2244. <https://doi.org/10.1021/acssynbio.8b00245>
109. Tizei, P. A. G., Harris, E., Withanage, S., Renders, M., & Pinheiro, V. B. (2021). A novel framework for engineering protein loops exploring length and compositional variation. *Scientific Reports*, 11(1), 9134. <https://doi.org/10.1038/s41598-021-88708-4>
110. Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes. *Plos One*, 4(5), e5553. <https://doi.org/10.1371/journal.pone.0005553>
111. Kalendar, R., Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., & Lohman, G. J. S. (2020). Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *Plos One*, 15(9), e0238592. <https://doi.org/10.1371/journal.pone.0238592>

112. Marquez, P. H., Koch, M., Kestemont, D., Arangundy-Franklin, S., & Pinheiro, V. B. (2019). Antha-guided automation of Darwin Assembly for the construction of bespoke gene libraries. *BioRxiv*, 2019.2012.2017.879486. <https://doi.org/10.1101/2019.12.17.879486>
113. Tóth-Petróczy, Á., & Tawfik, D. S. (2014). Hopeful (protein InDel) monsters? *Structure (London, England)*, 22(6), 803–804. <https://doi.org/10.1016/j.str.2014.05.013>
114. Emond, S., Mondon, P., Pizzut-Serin, S., Douchy, L., Crozet, F., Bouayadi, K., Kharrat, H., Potocki-Véronèse, G., Monsan, P., & Remaud-Simeon, M. (2008). A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. *Protein Engineering, Design and Selection*, 21(4), 267–274. <https://doi.org/10.1093/protein/gzn004>
115. Kashiwagi, K., Isogai, Y., Nishiguchi, K., & Shiba, K. (2006). Frame shuffling: A novel method for in vitro protein evolution. *Protein Engineering, Design & Selection*, 19(3), 135–140. <https://doi.org/10.1093/protein/gzj008>
116. Fujii, R., Kitaoka, M., & Hayashi, K. (2014). Random insertional-deletional strand exchange mutagenesis (RAISE): A simple method for generating random insertion and deletion mutations. *Methods in Molecular Biology*, 1179, 151–158. https://doi.org/10.1007/978-1-4939-1053-3_10
117. Hida, K., Won, S. Y., Di Pasquale, G., Hanes, J., Chiorini, J. A., & Ostermeier, M. (2010). Sites in the AAV5 capsid tolerant to deletions and tandem duplications. *Archives of Biochemistry and Biophysics*, 496(1), 1–8. <https://doi.org/10.1016/j.abb.2010.01.009>
118. Pikkemaat, M. G., & Janssen, D. B. (2002). Generating segmental mutations in haloalkane dehalogenase: A novel part in the directed evolution toolbox. *Nucleic Acids Research*, 30(8), 35e–35. <https://doi.org/10.1093/nar/30.8.e35>
119. Kipnis, Y., Dellus-Gur, E., & Tawfik, D. S. (2012). TRINS: A method for gene modification by randomized tandem repeat insertions. *Protein Engineering, Design & Selection*, 25(9), 437–444. <https://doi.org/10.1093/protein/gzs023>
120. Morelli, A., Cabezas, Y., Mills, L. J., & Seelig, B. (2017). Extensive libraries of gene truncation variants generated by in vitro transposition. *Nucleic Acids Research*, 45(10), e78. <https://doi.org/10.1093/nar/gkx030>
121. Osuna, J. (2004). Protein evolution by codon-based random deletions. *Nucleic Acids Research*, 32(17), e136–e136. <https://doi.org/10.1093/nar/gnh135>
122. Jones, D. D. (2005). Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 -lactamase to an amino acid deletion. *Nucleic Acids Research*, 33(9), e80–e80. <https://doi.org/10.1093/nar/gni077>
123. Liu, S. S., Wei, X., Ji, Q., Xin, X., Jiang, B., & Liu, J. (2016). A facile and efficient transposon mutagenesis method for generation of multi-codon deletions in protein sequences. *Journal of Biotechnology*, 227, 27–34. <https://doi.org/10.1016/j.jbiotec.2016.03.038>
124. Hallet, B., Sherratt, D. J., & Hayes, F. (1997). Pentapeptide scanning mutagenesis: Random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Research*, 25(9), 1866–1867. <https://doi.org/10.1093/nar/25.9.1866>
125. Murakami, H., Hohsaka, T., & Sisido, M. (2002). Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nature Biotechnology*, 20(1), 76–81. <https://doi.org/10.1038/nbt0102-76>
126. Li, G., Dong, Y., & Reetz, M. T. (2019). Can machine learning revolutionize directed evolution of selective enzymes? *Advanced Synthesis & Catalysis*, 361, 2377–2386. <https://doi.org/10.1002/adsc.201900149>
127. Ebert, M. C., & Pelletier, J. N. (2017). Computational tools for enzyme improvement: Why everyone can – and should – use them. *Current Opinion in Chemical Biology*, 37, 89–96. <https://doi.org/10.1016/j.cbpa.2017.01.021>
128. Mazurenko, S., Prokop, Z., & Damborsky, J. (2019). Machine learning in enzyme engineering. *ACS Catalysis*, 10(2), 1210–1223. <https://doi.org/10.1021/acscatal.9b04321>

How to cite this article: Alejaldre, L., Pelletier, J. N., & Quaglia, D. (2021). Title: Methods for enzyme library creation: Which one will you choose?. *BioEssays*, e2100052. <https://doi.org/10.1002/bies.202100052>