Molecular Biotechnology (2018) 60:946–974
https://doi.org/10.1007/s12033-018-0122-3

REVIEW

Tailoring Proteins to Re-Evolve Nature: A Short Review

Angelica Jimenez-Rosales[1] · **Miriam V. Flores-Merino**[1]

Published online: 27 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
Proteins are key biomolecules for most biological processes, their function is related to their conformation that is also dictated by their sequence of amino acids. Through evolution, nature has produced an immense variety of enzymatic tools of high efficiency and selectivity, and thanks to the understanding of the molecular basis of life and the technological advances, scientists have learned to introduce mutations and select mutant enzymes, to optimize and control their molecular fitness characteristics mainly for industrial, medical and environmental applications. The relationship between protein structure and enzymatic functionality is essential, and there are various experimental and instrumental techniques for unravelling the molecular changes, activities and specificities. Protein engineering applies computational tools, in hand with experimental tools for mutations, like directed evolution and rational design, along with screening methods to obtain protein variations with the desired properties under a short time frame. With innovations in technology, it is possible to fine tune properties in proteins and reach new frontiers in their applications. The present review will briefly discuss these points and methods, with a glimpse on their strengths and pitfalls, while giving an overview of the versatility of synthetic proteins and their huge potential for biotechnological and biomedical fields.

Keywords Directed evolution · Enzyme modification · Protein engineering · Site-directed mutagenesis · Random mutagenesis · Biocatalysis

Introduction

Enzymes are protein catalysts widely distributed in all biochemical processes that range from generation of energy, the activation of other proteins for performing their functions or inactivation of harmful molecules. Enzymes are highly selective since they accept a single substrate among a great variety of small biomolecules and large biopolymers to execute their reactions [1, 2].

Naturally, errors may arise in the codification of the sequence of amino acids of an enzyme, which results in a regular tendency for mutations to occur over an indefinite period of time, consequently randomly new catalytic activities arise, allowing for protein evolution. The mutability of these molecules was a feature that could be exploited to positively transform our daily lives, and until the mid-twentieth century, thanks to the availability of novel experimental techniques, scientists have used this evolutionary approach to explore and engineer new enzymes with a novel function at a laboratory-scale, this way artificially accelerating the evolutionary process of a single enzyme from millions of years to a few weeks [3]. This involves selecting the enzyme of interest (EOI) with a specific natural activity and to optimize its amino acid sequence through different methods to create a library of enzyme-variants for function screening and selection [4, 5].

Important developments in instrumentation, induced-mutation and genome-editing techniques, algorithms and automatized processes for catalysis screening, but the limitations in this area continue to focus on understanding the structure–function of enzymes, and on efficiently predicting the changes of these two components after modification of the protein structure [6, 7]. Nonetheless, the progress in this area has been substantial enabling the possibility to tune a given enzyme for a specific application, thus lessening the

✉ Angelica Jimenez-Rosales
ajimenez.qfb@gmail.com;
ajimenez@alumni.manchester.ac.uk

Miriam V. Flores-Merino
mvfloresm@uaemex.mx

[1] Center for Research in Medical Sciences (CICMED), Autonomous University of the State of Mexico, 205 Venustiano Carranza, Universidad C.P. 50130 Toluca, Mexico

boundaries in biotechnology to solve big challenges such as bioremediation; the production of new drugs, biocatalysts and biofuels; and innovating industrial processes into highly effective, low cost and environmentally friendly processes, among others [8–10]. Some ambitious trends are evolving entire biochemical pathways or genomes through genetic circuit design to create whole cell biocatalysts for advanced applications [11, 12].

The purpose of this review is to present the relationship between enzymatic structure and function, and how the methods of mutagenesis, protein design and selection are used to manipulate protein structure and improve the function, stability or specificity of an enzyme.

The Structural Balance Between Selectivity and Promiscuity in the Activity of an Enzyme

In biocatalysis, the induced fit model is the most accepted model of substrate binding to the active site of an enzyme to form the enzyme-substrate complex. It proposes that initially when both the enzyme and substrate bind, their interactions are weak, but soon, a conformational change in their structure occurs and their interactions grow stronger to allow the catalytic reaction. After which, the product is released, and enzyme will return to its native structure [13].

In this way, the three-dimensional structure of the enzyme-binding pocket and of the substrate pair with each other so that specific interactions occur, and the transition state is achieved faster, these features determine the rate of reaction and specificity of the enzyme. However, enzymes are flexible enough to alter their structural folding, and under physiological conditions, have more than one conformational state which affects the position of amino acids near the active site, altering the ability of the enzyme to selectively bind its substrate [14].

The function, specificity and stability of an enzyme relies on both intra-molecular (pi stackings, salt bridges, hydrogen bonds and hydrophobic interactions occurring within the molecule) and inter-molecular (occurring with solvents and their solutes such as ions and cofactors) interactions, which are fundamental for its structure [15]. Therefore, if an enzyme is in an environment that affects any essential interactions, it will change its properties.

Protein engineering of an enzyme is intended to modify its structure to improve function, stability or selectivity; as the designed mutations should introduce molecular and structural changes to form novel molecular interactions or provide them with new biochemical features to promote the desired catalysis. This is possible because, even if the current enzymes have evolved to specialize at a given biochemical reaction, often their selectivity is not absolute and show a minor activity to catalyse other reactions, and this alternative activity can be adjusted by mutation and selection, this way enzymes can be evolved artificially to acquire useful properties via the optimisation of their promiscuous activities [16] (Table 1). Experimentally, some mutations that partially reduce or increase the activity affect k_{cat} (catalytic rate constant) but the K_M (Michaelis constant) value remains, meaning that the enzyme-substrate complex had the same affinity but the catalytic efficiency is altered [17–19].

In addition, protein engineering improves our understanding of proteins and their function, as enzymatic screening with in vitro selection of mutants gives further insight about the structure, catalytic mechanism and activity of the new enzyme variations [34–37]; allowing to design strategies for predictions about the amino acids that are important to catalysis, or to adjust protein structure, and reduce constraining effects.

As should be expected, obtaining a novel specific and efficient engineered enzyme depends on the appropriate interaction of various biological, chemical and physical conditions, which are closely interconnected, and any modification may disrupt the balance between them, bringing serious alterations to the fundamental properties of the enzyme [38]. For example, in humans many natural genetic defects with single nucleotide polymorphisms are characterized by having problems in enzyme production or stability by aggregation, rather than the actual function of the protein. Thus, changes in enzymatic function rely in great proportion on stability, which depends on both expression and structural conformation [39].

Through protein engineering is possible to modify proteins to increase their applications. For one example is the green fluorescent protein, widely used as a reporter protein, for which through mutagenesis was possible to mutate specific residues in the chromophore of the protein to change the wavelength of energy emission, thus, modifying the fluorescent colour observed from green to variations of either blue, cyan o yellow, allowing for multifluorocrome-labelling methods [40, 41]. Another example is with an engineered transaminase for production of sitagliptin (anti-hyperglycemic drug) designed to substitute a rhodium-catalysed asymmetric enamine hydrogenation, the enzyme was modified for an efficient transamination of prositagliptin ketone, a compound for which initially it had no activity [42]. Computational tools and experimental protein-evolving techniques applied in the design of protein assemblies for molecule encapsulation, have allowed the production of libraries of synthetic capsid for RNA packaging which protects it from degradation from nucleases and extended circulation in vivo [43]. Considering potential vaccines, neutralizing antibodies of HIV-1 are produced by nearly 10% of patients with the disease, and identification of antibodies that interact with specific envelope (Env) epitopes of HIV has been studied to explore novel vaccine anti-viral antibodies through protein engineering to promote a wider range of

Table 1 Some examples of mutational modifications on proteins to change their activity, stability or selectivity for novel applications or understanding of its catalysis

Protein	Purpose of mutation	Technique	Result	Refs.
Dialkylglycine decarboxylase (DGD)	To change the substrate specificity of DGD from 2-aminoisobutyrate (AIB) to 1-aminocyclohexane-1-carboxylate (AC6C)	Directed evolution	Mutations N12D, R85H, N96S, N203S and S306F increase ~ twofold k_{cat} and ~ 2.5-fold decrease in K_M and, giving a modest ~ fivefold increase in k_{cat}/K_M for AC6C and a decrease in k_{cat}/K_M for AIB. Mutation S306F has the greater effect on catalysis	[20]
Dalcochinase	To increase the catalytic efficiency of dalcochinase towards a broader range of glycone substrates	Site-directed mutagenesis and computational analysis	Mutant F196H increased catalytic efficiency toward various glycoside substrates tested, with the most improved catalytic efficiency for hydrolysis of *p*-nitrophenyl β-D-mannoside of ~ threefold increase	[21]
DNA polymerase η (Pol η)	To understand the molecular determinants that govern human Pol η function in DNA polymerase, lesion bypass and fidelity activity	Site-directed mutagenesis	In polymerase activity, the naturally occurring isoforms of pol η tested (M14V, R81C, and E82D) and mutant Q38A displayed increased overall activity. Key residues for polymerase fidelity are Q38, Y52 and R61	[22]
Thermostable β-glucosidase 1Afrom *Thermotoga neapolitana* (Tn-Bgl1A)	To increase hydrolysis efficiency of Tn-Bgl1A of glucosylated flavonoids substrates to their aglycones derivates, specially quercitin-3-glucoside (Q3), using *para*-nitrophenyl-β-D-glucopyranoside (*p*NPGlc) as a model substrate	Site-directed mutagenesis and computational analysis	Mutations in residue 222 decreased ~ 2 to 3 times their conversion of *p*NPGIc. While mutations in residue 2221 increased up to ~ 3.5 times their conversion of *p*NPGIc compared to the WT	[23]
Bacterial serine protease Nattokinase (NK)	To increase oxidative stability of NK	Site-directed mutagenesis	Mutant M222A had an increased oxidative stability, as it was inactivated by more than 1 M H_2O_2 in comparison with WT that was inactivated by 0.1 M of H_2O_2	[24]
Chloramphenicol acetyltransferase (CAT) of *Staphylococcus aureus*	To increase the thermostability of CAT	Thermoadaptation-directed enzyme evolution	Protein produced in a heterologous error-prone thermophile derived from *Geobacillus kaustophilus* HTA426. Mutant A138T had a higher thermostability than WT with comparable activities	[25]
Baeyer–Villiger monooxygenase (BVMO)	To develop a BVMO system with high expression, activity and stability for the transformation of ricinolei acid into the ester Z)-11-(heptanoyloxy)undec-9-enoic acid)	Directed evolution	BVMO from *Pseudomonas putida* KT2240 was fused with a polyionic tag. This protein and alcohol dehydrogenase of *Micrococcus luteus* were expressed with a strong synthetic promoter in transformed *E. coli* that was able to perform all the product transformation at a laboratory-scale (85% and 17.2 mM/h) and 5 L scale (72% and 15.7 mM/h)	[26]

Table 1 (continued)

Protein	Purpose of mutation	Technique	Result	Refs.
Horseradish peroxidase (HRP)	To engineer a new yeast *Pichia pastoris* strain for production of HRP with reduced hyper-glycosylation	Combination for protein and strain engineering	Enzyme variant 4/8 HRP (N13D, N57S, N255D and N268D) shows considerable catalytic activity and thermal stability and is less glycosylated	[27]
β-Glucosidases (BGL1) from *Aspergillus niger*	To engineer a *Saccharomyces cerevisiae* strain that expresses a functional cellulase system that simultaneously hydrolyses biomass while fermenting hydrolysate	Directed evolution, site-directed mutagenesis and computational analysis	Mutants were obtained through heterologous expression in *S. cerevisiae* of enzyme-variants of BGL1 from *Aspergillus niger*. Y305C reduced the inhibitory effect of high substrate concentrations towards the hydrolytic reaction. Mutant Q140L, had increased hydrolytic activity of 156% compared to BGL1 at 2 mM pNPG. The double mutant Q140L and Y305C had slightly higher K_M and K_i values compared to Y305C	[28]
Cellodextrin transporter 2 (CDT2) from *Neurospora crassa*	To improve cellobiose fermentation with CDT2 under anaerobic conditions	Directed evolution	Mutant CDT2 HHT (triple mutant Q207H, N311H and I505T) had an increased cellobiose uptake activity by ~2.2-fold. During high cell density fermentation in anaerobic conditions the increase was of 4.0- to 4.4-fold in the cellobiose consumption rate and ethanol productivity respectively	[29]
Flavin-dependent halogenase RebH	To change the substrate selectivity of RebH to preferentially chlorinate tryptamine rather than Trp	Protein structure-guided mutagenesis approach	Mutant Y455W was transformed into the alkaloid-producing plant Madagascar periwinkle (*Catharanthus roseus*). The organism showed de novo production of the halogenated tryptamine: 12-chloro-19,20-dihydroakuammicine	[30]
Protein arginine methyltransferase 7 (PRMT7)	To study the exclusive product selectivity of mono-methylarginine (MMA) products to di-methylarginine (DMA) of PRMT7	Site-directed mutagenesis	Double mutant containing, E181D and NQ29A changes the enzymatic product specificity and allows the symmetric DMA product	[31]
RNA-binding protein TDP-43	To study mutations on TDP-43 that causes changes if the protein found in amyotrophic lateral sclerosis	Site-directed mutagenesis	Mutant D169G, had higher thermal stability than WT and was easily cleaved by caspase 3, producing increased levels of the pathogenic C-terminal 35 kD fragments (TDP-35). D169G mutation changes the conformation to a β turn, increasing the hydrophobic interactions that result in thermal stability	[32]

Table 1 (continued)

Protein	Purpose of mutation	Technique	Result	Refs.
Kunitz protease domain in human amyloid precursor protein inhibitor (APPI)	To develop a biological inhibitor for mesotrypsin by engineering APPI for increased proteolytic stability, stronger binding affinity and improved selectivity for mesotrypsin inhibition	Directed evolution	APPI triple mutant (M17G, I18F and F34V) had a strong mesotrypsin inhibition constant (K_i) of 89 pM. The variant displays 1459-fold improved affinity, up to 350,000-fold greater specificity and 83-fold improved proteolytic stability compared with WT. Also, the APPI triple mutant acts as a functional inhibitor in cell-based models of mesotrypsin-dependent prostate cancer cellular invasiveness	[33]

AC6C, 1-aminocyclohexane-1-carboxylate; AIB, 2-aminoisobutyrate; APPI, amyloid precursor protein inhibitor; BGL1, β-glucosidases; BVMO, Baeyer–Villiger monooxygenase; CAT, chloramphenicol acetyltransferase; CDT2, cellodextrin transporter 2; DGD, dialkylglycine decarboxylase; DMA, di-methylarginine; ee, enantiomeric excess; HRP, horseradish peroxidase; MMA, mono-methylarginine; NK, Nattokinase; *p*NPGlc, *para*-nitrophenyl-β-D-glucopyranoside; Pol *η*, DNA polymerase *η*; PRMT7, protein arginine methyltransferase 7; Q3, quercitin-3-glucoside; Tn-Bgl1A, thermostable β-glucosidase 1A from *Thermotoga neapolitana)*

neutralization of HIV-1 strains which could be potentially used in passive immunization or treatment [44]. These examples show a few useful applications and similar approaches may also be applied to any other protein or enzyme, thereby potentially expanding their applicability and versatility.

Approaches for Redesigning a Protein

Changing the stability, selectivity or activity of an enzyme is no easy task, even predicting changes on function by a single amino acid mutation is complicated as often there is no structure–function information [45] and any mutation, especially of conserved residues, and may lead to folding instability [46], nonetheless, there are methods to approach protein modification.

Rational protein design is a method that requires full comprehension of the protein structure and its catalytic mechanism, to systematically use site-directed mutagenesis to alter the active site and change the EOI in the desired way [19, 45]. Therefore, the structure, amino acids, charges and distances of the enzyme should be deeply analysed, and often it is difficult to predict the effect of the mutation [4, 8, 19, 47].

Directed evolution is a second method which does not require understanding of the structure of the enzyme and relies on the natural selection of proteins after several rounds of random mutagenesis [48]. Thus, based on unexpected changes in the genetic material and an adequate method for selection the EOI is isolated; however, the main disadvantage is the need of a high-throughput assay to screen for activity [49–51].

The iterative saturation mutagenesis (ISM) scheme, is a systematic probing of only a segment of the EOI in order to minimize the screening effort by reducing the number of residues of a given randomization site, and this has proved to be effective in improving the catalytic features and binding affinity of proteins [52–54].

The data-driven design or semi-rational design uses the information available to select mutations while allowing unpredictable substitutions that produce mayor changes. This method greatly relies on computational modelling and bioinformatics [55].

In research, these methods complement each other and can be used in parallel towards the evolution of an EOI.

Computational Analysis for Protein Structure Prediction

Experimental strategies may lead to large libraries with a time consuming and costly screening, and unfortunately, most mutants will have no better features than the WT enzyme. To overcome these big challenges, computational

analysis came as an option to design targeted smaller libraries based on the data available.

There are numerous computational tools available, but in a simplistic description, these algorithms are based on different principles, a general approach to describe them may consider the "*ab initio* methods" whose foundations are chemical and physical principles to propose the minimal energetical cost in the protein folding and structure-affinity analysis on several types of substrates. The "comparative methods" that may include threading recognition, fragment based [56] and homology modelling, use similarities and repeated features among families of proteins to model the structure of ancient, current and novel proteins. The "empirical methods" that use heuristic and profiling strategies to solve a structure when the search space is immense make an exhaustive exploration impossible. The evaluation of the model is the most important aspect because it must stick to the scientific principles of chemical structures and must present the relevant biological features as close as possible with reality. Most of times, this last aspect is the hardest to assess and that produces greater uncertainty [9, 57–62]. Applications of these methods along with other experimental techniques have aided in this "enzyme reconstruction and resurrection" [63], which has brought powerful catalysts that are active and stable at different physicochemical conditions from the current existing enzymes [64].

Some examples of these computational methods are Phyre, SWISS-MODEL, FoldX, Modeller, ROBETTA, PoPMuSiC [65–67], MOSST (Mutagenesis Objective Search and Selection Tool) [45, 68], SCHEMA [69–71], PINGU (PredIcting eNzyme catalytic residues usinG seqUence information) [72] and THEMATICS (Theoretical microscopic titration curves) [73, 74].

Some limitations for computer algorithms include that the processing of data takes a long time, but currently there are megaprojects such as Folding@home, by which people around the world can connect to a server and contribute to the processing of structures [75].

Experimental methods for enzyme design and selection

Protein structure has a high level of complexity within its inter-molecular and dynamic forces, therefore in the analysis of its structure is important to consider its folding stages, structural rearrangements, substrate binding and catalytic activity, among other specific features [76]. Any enzyme to obtain its final functional conformation must be folded into its definitive 3D structure, which is a process that starts as early as necessary, and in which several chaperone proteins, cofactors and metal ions are involved to assist in the folding [77–79]. Also, conformational transitions are crucial during catalysis, and many of these transitions involve disordered regions in the enzyme, which can exhibit various rearrangements that are essential for activity. It is important to note that disordered regions are different from flexible regions that have various conformations but do not intervene in the catalytic activity of the protein, and the distinction between cannot be made on the sole basis of a three-state secondary structure [76]. Therefore, modification of an EOI requires to observe its functional cycle and structure, as it is a constant-dynamic and environmental-interactive macromolecule.

The combination of various computational methods can enhance the accuracy of predictions, however, experimental assessment of the modified EOI is required and this requires several methods and approaches to observe if the desired results were achieved. For this purpose, protein engineering uses molecular biology techniques and qualitative-quantitative analytical techniques.

DNA editing methods

DNA cloning in molecular biology is mainly used to isolate a gene from the rest of the cell genome to modify it and propagate it in the same or in a different species, later the cloned gene can be introduced into an expression vector for the production of mutated proteins. Currently, these techniques are a routing practice and there are mainly two approaches, for one, a gene is isolated from genomic DNA by selectively cutting it with restriction enzymes, and the gene (or genes) DNA fragment is directionally inserted and ligated though DNA ligase into a vector, which can be a viral vector, a plasmid or prokaryotic/eukaryotic artificial chromosomes. Then, the small recombinant DNA is introduced into competent cells to replicate and produce the intended protein, or to isolate the plasmid and generate a library of mutants. The second approach is the generation of cDNA from mRNA transcripts by using reverse transcriptase, which represents a tremendous advantage for eukaryotic genes that are edited from a pre-mRNA into a mature mRNA. After this step, the cDNA is directionally cloned into the vector for later the transformation of competent cells or to as in the previous approach, to generate a library of mutants [80].

There is a variety of cloning methods, that range from restriction and ligase enzymes-dependent, to ligation-independent, to PCR-based, to recombination-based cloning, all intended to increase ease of the procedure, efficiency and yield. These strategies can be selected according to the length, repetitions and complexity of the sequence, the number of genes, the expression system and the tags to use. Selecting the vector is of equal importance, as the production of functional enzyme, especially some enzymatic complexes, carries another challenge due to the folding, solubility, assembly and co-expression of specific subunits.

Approaches for this include using multiple vectors fused with a single gene and selection marker to introduce genes into host cells. Another is poly-cistronic constructs for the expression of multiple genes from a single vector and under the control of a single promoter, but each gene has its ribosomal binding site. Finally, using a single vector but each gene is regulated individually by a separate, equal or different, promoter. One must consider that design of the cloning strategy should be flexible for a combinatorial approach that is often required for multi-gene constructs and PCR-based methods.

However, sufficient amount of soluble protein relies not only on the insertion of the gene into a particular vector, but also depends on the selected of expression system, the conditions used and vector topology [81].

Complications may arise with genes that contain internal restriction sites present in the cloning site, during expression constructs with poor compatibility between the cloning sites and vectors, when prokaryotic transcription and translation systems produce an enzyme with bad performance, when the unintentional introduction of additional nucleotides into the coding sequence leads to the addition of non-active extra amino acids in the EOI that may hinder activity or in repetitive protein coding genes. Fortunately, solutions to these problems have been developed such as multiple-host vectors that permit protein expression of the same construct in bacteria, insects and mammalian cells, cloning protocols for templates to recombine multiple fragments with no sequence homology and powerful recombinant vector systems for multiprotein assemblies [82], among others. Therefore, with several cloning strategies available, more than one option will be useful (Table 2).

Generation of mutations and recombination in the gene (or genes) of interest, is of utmost importance for protein engineering. Parent genetic template sequences, either created in the laboratory or from a natural origin, are modified into novel combinations of sequence information to generate mutant enzyme libraries, which are later screened for the desired function.

There are dozens of methods described that have been developed for DNA manipulation to create specific constructs in a broad range of organisms with the intention of introducing insertions, deletions, mutating either a single site or various sites randomly, etc. Generally, two different approaches for the generation of mutant libraries exist: asexual (non-recombinant) and sexual (recombinant).

For example, site-directed mutagenesis, is a very important asexual method in the study of structure–function relationships of genes and proteins. It a technique that uses two complementary synthetic oligonucleotides, which contain the mutation at the corresponding codon, to facilitate the introduction of a single point mutation in the gene (Fig. 1). Site-directed random mutagenesis (site-saturation), is a derived technique from the previous technique, for which the oligonucleotides used contain a degenerated codon encoding for all amino acids to mutate a selected position in the gene [77]. It targets a specific residue, thus mutants in the library tend to maintain the WT structure of the enzyme with high probabilities of functionality and there are several cases where it has been successful for alteration of substrate specificity or enantioselectivity, novel catalytic activity, enhanced stability, reduced immunogenicity, among many other applications. It is a PCR-based method with full amplification of the gene and vector, and to eliminate the original WT-DNA, DpnI is used, that as it is that an enzyme that can degrade methylated DNA. However, there are limitations with this method by using complementary oligonucleotides, one being that cannot introduce several mutations, and other that dimers may be produced [87, 88].

Combinatorial cassette mutagenesis is another asexual technique used for simultaneous saturation mutagenesis in multiple sites, as it is capable of introducing segments of random arrangements of DNA into a target sequence without adding other non-native sequence. It requires a DNA cassette with restriction sites available to insert the gene(s) of interest or the mutated sequences flanked by two restriction sites with the same cleavable motif as the cassette, for their ligation. If the restriction sites are not present in either the cassette or the sequence of interest, or both; they can be designed by site-mutagenesis, with primers codifying for a specific restriction site according to the enzyme to be used. This method can easily generate combinatorial libraries that after ligation to a s, the can be selected according to their functionality protein is carried out (Fig. 2) [4, 78].

Error-prone PCR (ep-PCR), is considered an asexual technique based on inaccurate copying of a sequence by DNA polymerase, which under specific experimental conditions or type of enzyme used, it may add 23 incorrect bases to each replicated DNA strand of the gene. Unlike site-directed random mutagenesis or cassette mutagenesis, it allows introduction of random mutations within a wide range of the target gene, simulating the natural process of natural random mutagenesis, thus, it is a widely used technique, but the biased occurrence of amino acids is an intrinsic drawback (Fig. 3).

DNA shuffling is a random recombination of DNA fragments from homologous genes into full-length chimerical gene. It simulates the process of natural recombination by digesting DNA sequences of homologous genes and combining the fragments by denaturation, annealing and elongation with a DNA polymerase (Fig. 4). Comparing DNA shuffling with ep-PCR, DNA shuffling can be applied to sequences > 1 kb, while ep-PCR does not allow fragments > 0.5–1.0 kb. However, both techniques can be used in a pool of unknown sequence and share a similar mutagenesis rate. DNA shuffling with WT-DNA removes neutral

Table 2 Typical cloning strategies commercially available

Technique	Name	Description	Advantages
Traditional cloning	Restriction enzyme and ligation cloning	Double stranded DNA is cut into fragments with restriction enzymes at a specific site leaving either sticky or blunt ends and are ligated. If a single restriction enzyme is used, then non-directional cloning occurs, and clones need to be screened for gene orientation. If two different restriction enzymes are used, then two different ends are produced, and the clones have the correct gene orientation	It is cheap and has hundreds of enzymes available. It has no mayor problems for directional cloning
Recombinational cloning	TOPO® cloning	It uses topoisomerase I to cleave and re-join a specific hotspot sequence in the vector, which is a 3′-tymine. After addition of the DNA fragment, insertion and ligation occurs by linking of topoisomerase I. TOPO TA cloning is a variation when DNA inserts are prepared using PCR and Taq polymerase that leaves an adenosine 3′ end overhang, and thus the vectors a 5′ thymine overhang to pair with it	Simple procedure with high yield, fast and accuracy. It does not require restriction endonuclease at the cloning step, fewer steps are required. Some vectors already have topoisomerase I covalently attached to the 3′-phosphate group at each end
Recombinational cloning	Gateway™ cloning	DNA fragments are cloned into a donor plasmid through specific recombination sites, then this "entry clone" can be easily sub-cloned into different compatible "destination vectors" to carry out different molecular biology techniques depending on the objective	Allows inserting different promoters for your gene, to fuse your protein of interest to different tags, or putting your gene under different selection cassettes. Reaction is directional and specific
Seamless cloning	Isothermal assembly reaction or Gibson Assembly	Uses 5′-exonuclease to digest the 5′ end of the dsDNA leaving 3′ single stranded sticky ends that will find a match and anneal. Then polymerase will complete the gaps to create a complementary DNA strand. Finally, a ligase fuses the backbone phosphate bonds to produce a single dsDNA	Any DNA fragment can be used to be ligated into any plasmid. The user has total control of the cloning since it does not rely on restriction sites. Various DNA fragments can be joined in a single reaction
Seamless cloning	Golden Gate Assembly	Uses type IIS restriction endonucleases that cut dsDNA at a specified distance from their recognition sequence to leave overhanging sequences in the insert for ligation with the vector. The type IIS enzymes have different recognition sites for a 5′ and for 3′ overhanging ends	One-pot process with a single restriction enzyme. The cloning follows directionality thanks to the longer and discrete overhangs of the DNA. The recognition sequences are removed from the final product by the type IIS restriction enzyme, thus no "scar" sequence is retained
Ligation-Independent cloning	LIC	Uses T4 DNA polymerase in its 3′→5′ exonuclease activity to create 5′ single stranded overhangs in vector and DNA insert with the presence of one dNTP the exonuclease activity will occur until there is complementarity in the first residue	Complementary overhangs are long; therefore, it is very specific. No reaction for ligation is necessary and the product can be directly transformed and repaired in the bacteria

Table 2 (continued)

Technique	Name	Description	Advantages
Large oligonucleotide assembly	Yeast-mediated cloning and oligonucleotide stitching	Uses yeast to fuse various dsDNA fragments that share < 30 bp overlapping. It just requires the fragments of DNA to be joined and DNA custom oligos of 60–80 bp sharing 30–40 bp with the DNA fragments to fuse	Larger product sequences of DNA fragments, genes and genomes. Yeast can fill gaps of 160 pb. DNA pieces that do not share complementarity can be fused

Methods consulted from New England Biolabs, The European Molecular Biology Laboratory, Addgene [83–86]

mutations produced by repeated cycles of any mutagenesis strategy [89].

DNA shuffling was demonstrated by Stemmer in 1994 and since then, numerous in vitro recombination methods have been developed. Some of these methods for recombination and also methods for asexual mutations are described in Table 3.

Other types of methods involve chemical or physical random mutagenesis, which are based on the variety of chemical reagents and physical factors that are reported to mutagenize DNA by inducing alkylation's, deamination, pyrimidine dimers, base oxidation among others.

There are various perspectives for the advantages and disadvantages of each method. Considering a point of view from natural evolution that is a function of variation and selection, computational simulation studies of evolution of protein sequences and their structural classes, have demonstrated homologous recombination in the evolution of biological systems, which appears remarkably advantageous by combining valuable mutations that have arisen independently and may be synergistic, while simultaneously removing mutations that decrease the fitness of an organism. Unfortunately, all these methods generate large numbers of mutants, most of them non-functional, and therefore the screening is extensive and requires high-throughput methods that are costly and time consuming.

Using strategies to achieve enzyme modification for stability, activity and selectivity

Protein modification may implicate producing multiple variants as the number of amino acids and mutations is enormously vast. However, random mutant libraries of a few hundreds or thousands of mutants are enough to obtain a stable and active enzyme, as it is not necessary to exhaustively examine an EOI in each and every single position, but rather create a random mutagenesis library to get a statistically significant number of mutants with the desired features when comparing them to the WT protein. After the screening, a few variants may be chosen, and if necessary, deeper work on the selected mutations can be performed [100].

Site-directed mutagenesis is commonly used in rational protein design to manipulate the fine balance between structure, ligands and protein recognition to obtain new desirable features, and in some cases even transform the activity of the EOI. This process that naturally occurs in divergent evolution throughout species, and it is followed by functional evolution or neo-functionalization [101]. For instance, the membrane associate guanylate kinase (MAGUK) contains a GUK domain (GUK_{dom}) that is used to interact with the proteins of the cytoskeleton for adhesion and signal transduction.

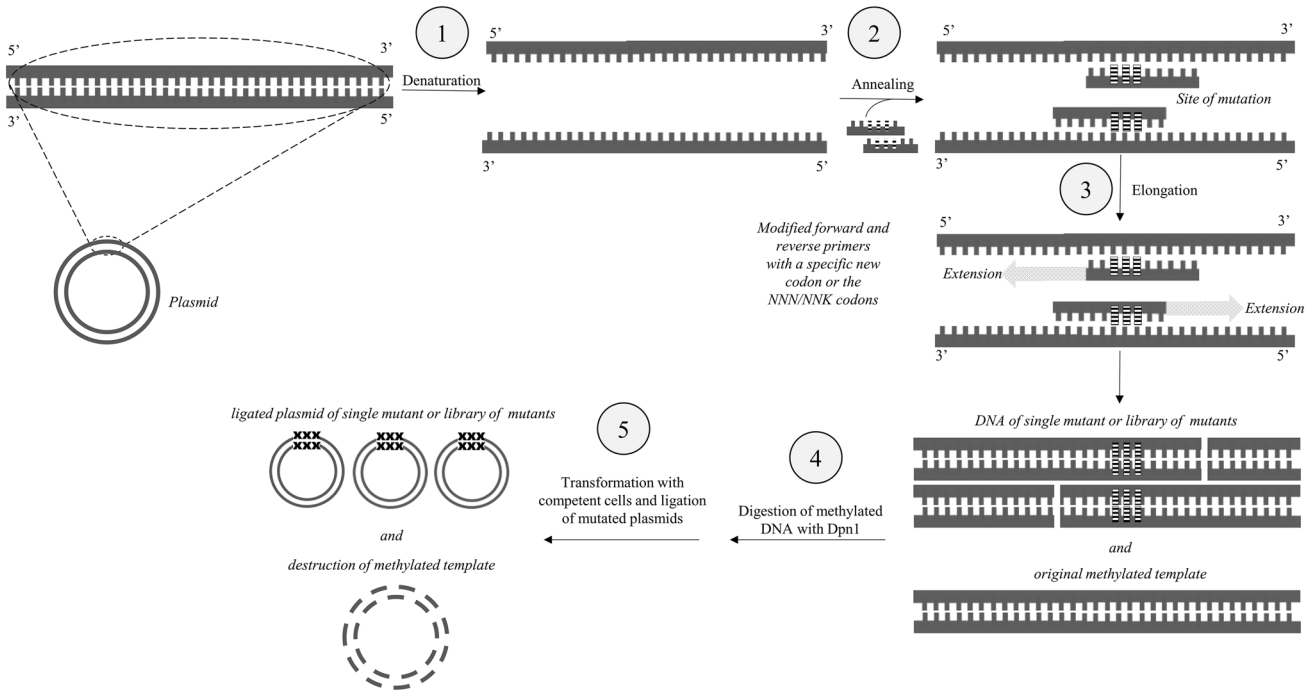

Fig. 1 Site-directed mutagenesis creates targeted changes in dsDNA plasmid by using designed oligonucleotide primers (~25 bp) in a regular PCR to amplify the full vector and confers the desired mutation with a selected amino acid. It can be used to change a single codon to another, to add or remove a small sequence of codons (~20 nucleotides). In site-saturation mutagenesis, the primers are degenerated at a designated codon (in diagram shown as striped nucleotides or an X), either NNN (where N is any nucleotide) or with a codon NNK (where K is either a T or a G). The original methylated template must be digested with DpnI before transformation. The primers yield a circular doubly-nicked plasmid that can be directly transformed into competent bacteria where it can be ligated in vivo into a circular DNA by bacteria or by an added kinase and ligase

However, other enzymes have the GUK catalytic domain (GUK_{enz}) that is used to bind ATP [101, 102]. Comparison of these two GUK domain sequences in *Drosophila* revealed that aside of a P-loop, the sole different residue between GUK_{enz} and GUK_{dom} is Ser-68, which is a Pro respectively, and it is key residue which marks the functional difference between a nucleotide kinase and a recognition domain [101].

Enzymes must be functional and stable in the cell, so their folding, aggregation and degradation are primary structural aspects. So, in directed evolution, the more stable the EOI is, the bigger the range it has for mutational changes and evolvability. In this aspect, mutations that increase stability can help to balance the disrupting effect of other mutations, unfortunately these modifications are permanent during the functional cycle of a protein and do not provide any further advantage. An alternative is using molecular chaperones, providing a temporal buffering that does not affect the production and concentration of the enzyme. Moreover, chaperones not only provide stability to some mutations, but also help with the folding of the protein [19, 103].

During interaction of enzyme and substrate, the availability and location of the binding site are central for activity. Therefore, orienting the EOI with a specific alignment through immobilization is a possibility to promote higher activity and carry out sequential synthesis in a single step by co-immobilizing different related enzymes [104]. Site-directed mutagenesis has facilitated to immobilize enzymes, and even in some cases, it has improved thermostability and stability with organic solvents [105]. Additionally, cell surface display, which is a technique that permits the expression of proteins fused to membrane-proteins of bacteria, phage or yeast; works similarly for immobilization and has allowed the development of vaccines, drug-delivery systems, new catalysts and bioremediation alternatives [8].

Screening for the enzyme of interest

Mutagenesis in the starting point in protein engineering, a suitable screening is key step in this process to obtain the variants of interest in short time and less labour-intensive. The strategy to implement it should consider correct production of the protein from the host, the efficiency of transformation, assay development and its reproducibility and equipment needed, among others. All techniques have advantages and disadvantages; thus, the idea is to follow a complementary approach to designing high-throughput strategies for protein purification and measuring activity (should

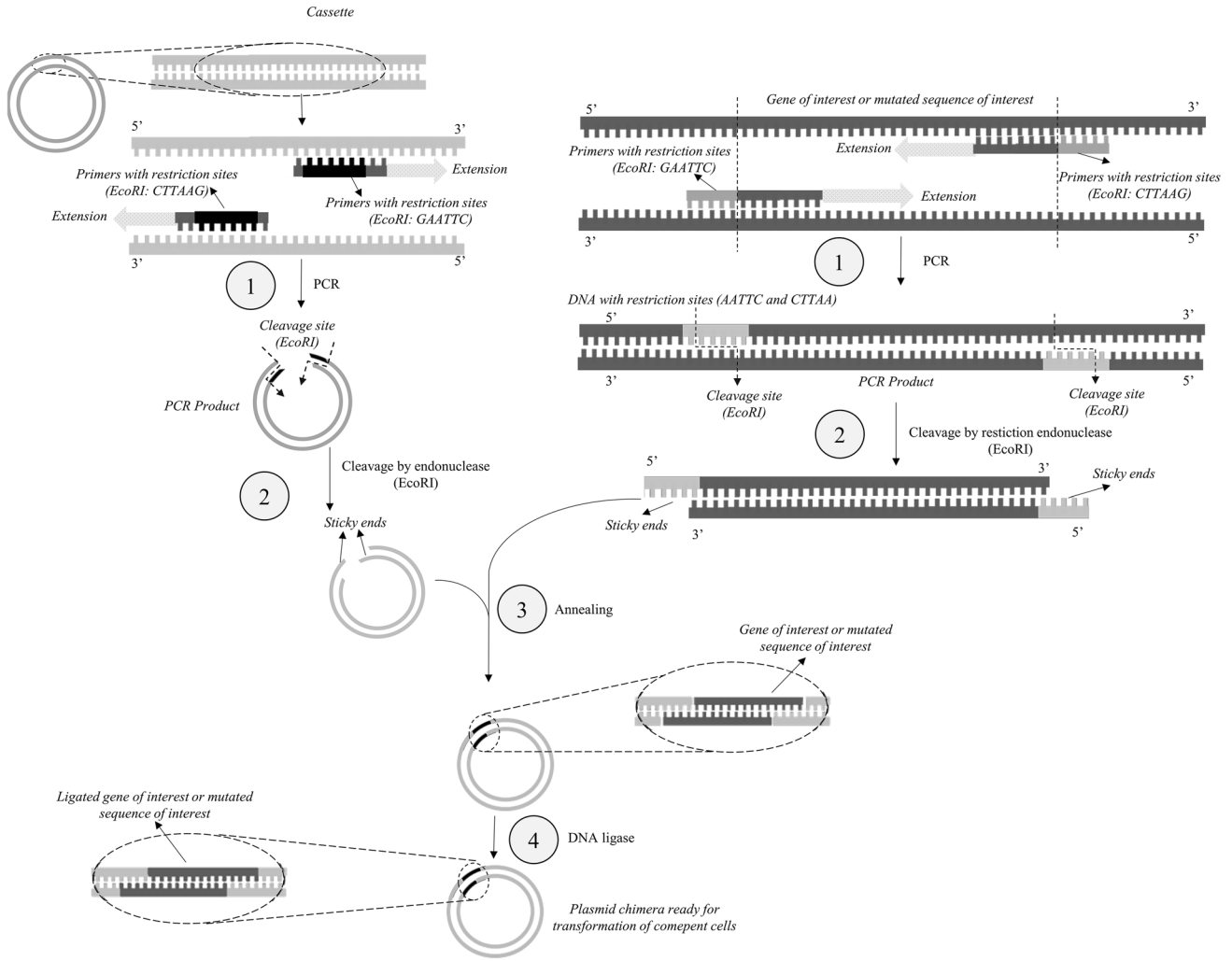

Fig. 2 Combinatorial cassette mutagenesis allows the insertion mutagenic oligodeoxynucleotide cassettes. It is performed by inserting the DNA sequence (dark grey) in the cassette (light grey) using the corresponding restriction sites (in diagram: *Eco*RI, black for cassette restriction sites and light grey for sequence restriction sites). In both cases the fragment of the gene to be inserted and the DNA where it will be ligated must have sticky ends so that they align, and latter can be covalently ligated with a DNA ligase. After ligation takes place, this construct can be used for cell transformation

be straightforward, inexpensive and fast) either with assays for broad substrate specificity assessment or for substrate profiling, to screen protein with the desired substrate.

It is best to have the protein as pure as possible as impurities might interfere with results and the screening of enzymatic activity will involve measuring any of the components of a reaction: the free enzyme, the substrate or the product. Each of these may be directly detected by specific labelling or if the molecule is not spectroscopically active, it may be indirectly detected by coupling the reaction to a reporter molecule. Thus, isotope labelling, fluorescent products, detection antibodies and tags, are highly desired in these experiments. If not available, there are other instrumental techniques that can be used to detect special products and to study their structure and molecular changes such as mass spectrometry or nuclear magnetic resonance, a brief description of these is presented in Table 4.

Selection of a direct or indirect detection method depends on the high sensitivity of the method of detection and the stability of the measured component is required. Detection techniques for data collection of a reaction may be limited by automatization, real time monitoring or requirements of sample processing. In indirect detection schemes, a similar rate of transformation and saturation are also important, so that the original reaction is not limited by the kinetics of the reporter reaction.

Determination of the K_m and k_{cat} are of main importance as they serve to standardize the assay and are critical for the assessment of activity and stability of the expression of constructs, determining of inhibitors, substrate binding,

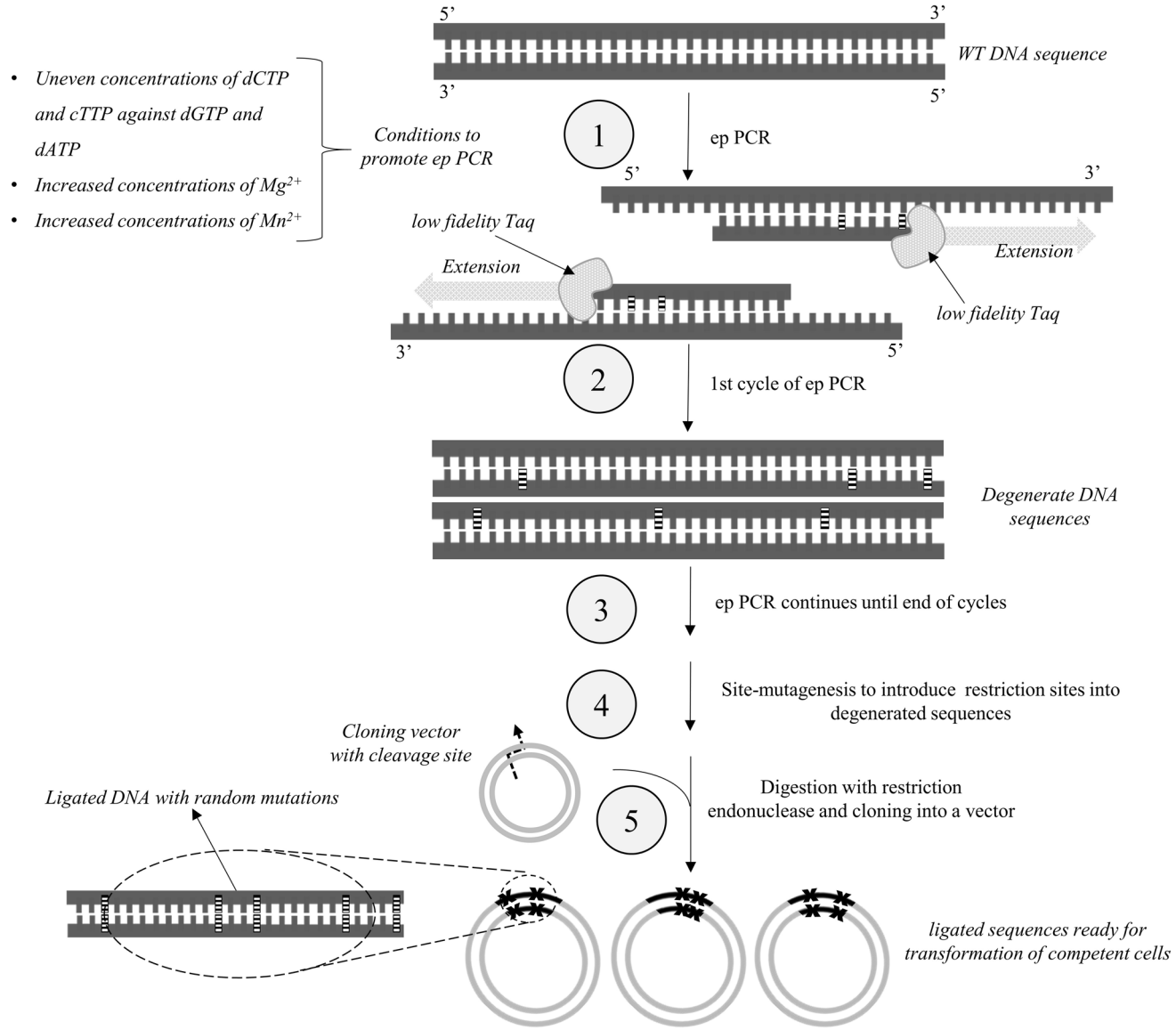

Fig. 3 Error-prone PCR (ep-PCR), it is a PCR that takes advantage of the low fidelity of Taq polymerase and other engineered polymerases, to insert randomly a mispair when performing DNA polymerization. Another method to promote mutations is creating non-ideal conditions of the reaction, such as spiking some of the dNTPs, either dCTP and dTTP with dGTP and dATP, also by increasing the concentration of $MgCl_2$ or $MnCl_2$. The sporadic mutated bases along the DNA segment (in diagram shown as striped nucleotides or an X) promote random mutations for a different codifying codon. After generation of DNA, the sequence can be subjected to site-directed mutagenesis, to introduce restriction sites for further ligation into a vector

compound screening and other factors that have significant consequences in a biochemical reaction [116, 117].

Experimental techniques for assay screening of the EOI are varied and use different conditions, but as general rules: (1) The concentration of the EOI depends on the method of choice, but it is suggested to do dilutions where it is in complete solution, and select the proper dilution according to the signal obtained in the instrument (thus minimizing the sample consumption and spent effort), (2) The tertiary structure of proteins is observed near the UV region where aromatic amino acids absorb [118], (3) Osmolytes, modulate their macromolecular properties according to their protonated states, which can lead to undesired results that do not normally occur in native conditions [119].

For protein stability, the difference of free energy ($\Delta G°$) between the folded and unfolded states is key in thermodynamic stability, however, just a minimum fraction of protein may be unfolded in native conditions, hence the difficulty in quantifying the folded-unfolded state equilibrium, and it is necessary to change the conditions to shift the equilibrium

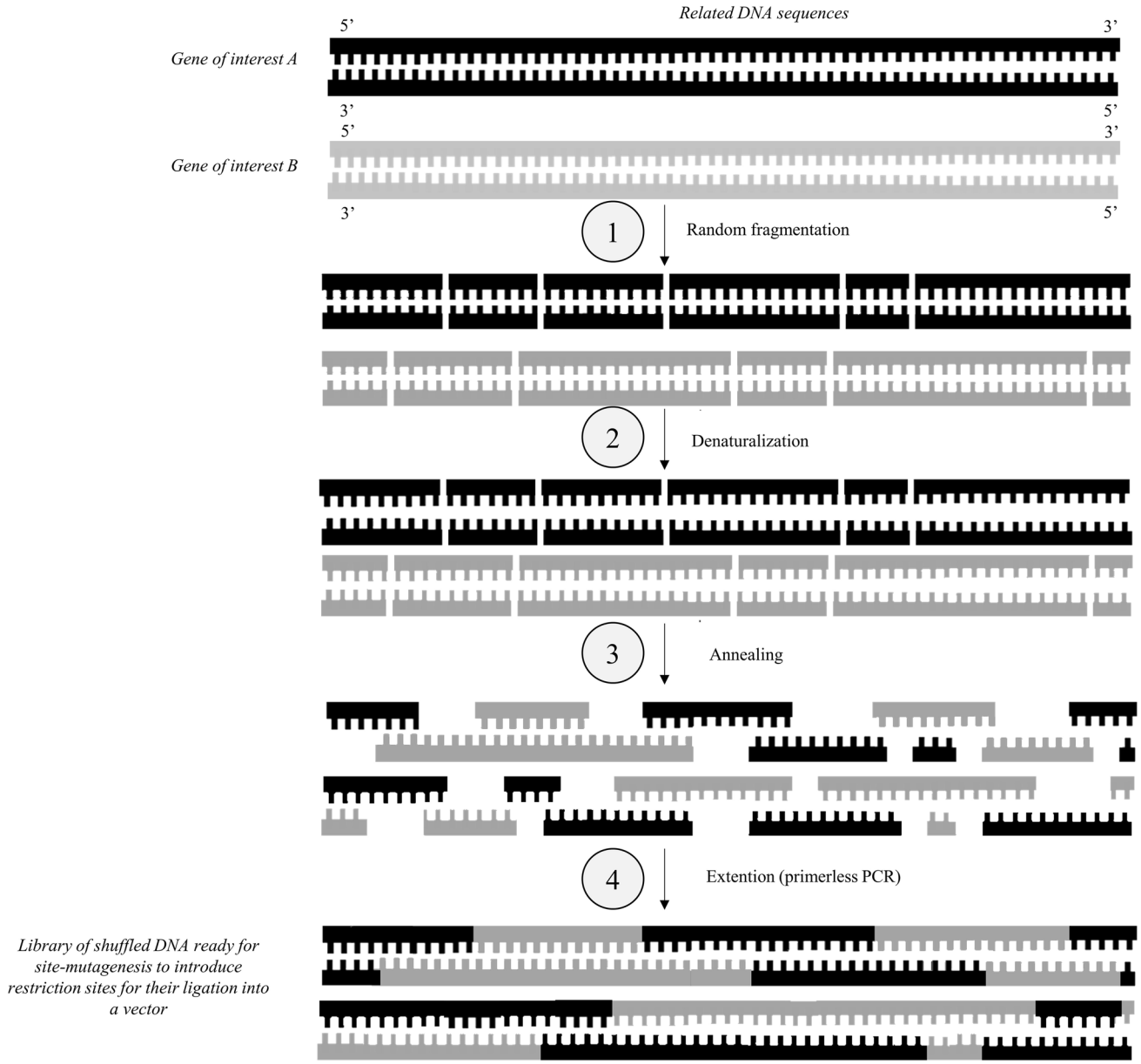

Fig. 4 DNA shuffling is a technique for the recombination of homologous gene sequences (gene 1 black and gene 2 grey). The genes are digested with DNAse to obtain random small fragments which are subjected to melting, annealing and extension in a PCR-like process with no added primers [90]. Finally, the sequences obtained can be cloned into a vector once the restriction sites are inserted in the sequence by a site-directed mutagenesis

[82, 119, 120]. The thermostability can be tested by measurement of the residual activity under heat exposition since most protein structures will loss activity as they denature and precipitate at high temperatures [119]. However, thermostability and even proteolytic resistance does not always relate with solubility and functionality, as there are cases where the protein variants are expressed as insoluble inclusion bodies [19].

Solubility is an essential feature but several times it is no easy task to assess it due to changes in viscosity, pH shifting, binding to surfaces or aggregation, among others [121]. Normally, the activity of the cell lysate is measured to identify the concentration at which the protein is soluble and active [119, 122]. Another method may be indirect assay with a reporter protein, where its activity is related with the solubility of the EOI. Typical examples of reporter proteins are the green fluorescent protein, β-galactosidase, dihydrofolate reductase or an antibiotic resistance protein [19].

Predictor methods for solubility have also been developed; unfortunately require specific conditions, as rely on a

Table 3 Typical methodologies to obtain mutations or recombination

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
Megaprimer and ligase-free PCR mutagenesis	An asexual technique for cloning a fragment of DNA with mutations (either random or site-specific) into a plasmid	The DNA segment that will be cloned is used as a megaprimer with mutations in its sequence that substitutes a homologous region in the template plasmid	Overcomes limitation of conventional ligations Novel methodologies use forward and reverse primers to produce the megaprimer and then with different annealing temperature amplify the full plasmid	Use a high-fidelity polymerase for insertion to minimize new mutations If the DNA to be inserted was prepared using Taq DNA polymerase, is important to remove the extra nucleotides at the 3′-end of amplified products to avoid unintended mutations	Preparation of DNA (megaprimer).to be inserted by the desired mutagenic technique PCR with the megaprimer and plasmid Dpn1-treatment Transformation of cells, in vivo repairing of nicks in DNA
Gene assembly mutagenesis	An asexual method to produce whole genes, plasmids and viral genomes with mutations through oligos	A set of oligos are designed to cover the length of a gene. Swapping the codons in the oligos to replace more than one amino acid allows the insertion of mutations into the reassembled gene	Mutations can be easily performed in a single step and targeting several sites by exchange of oligos. Prevents contamination from the unmutated template	Introduces non-targeted mutations at an elevated rate The assembling reaction in large assemblies produces a great variety of products of all types of sizes It may be necessary to clone the gene into a pSALect vector for in-frame selection	Design of the oligonucleotides of ~40 bases considering plus and minus strands Perform the PCR to assemble oligos Amplification of the assemble using the outermost primers to obtain a single product Cloning of PCR product
Mutator strains and mutator plasmids	An asexual method and involves the introduction of random mutations along the entire gene by transformed mutator strains or by using a mutator plasmid that temporarily converts normal bacteria into mutator strain	Mutator strains are organisms with deficient DNA repair genes, thus their offspring can rapidly evolve to present new genetic variations Mutator plasmids carries an inactive mutant of *mutD* gene, which prevents proof reading of DNA polymerase III, also this gene is sensitive to temperature, thus it is easy to control its expression	Mutator strains evolve rapidly to produce novel proteins Since the process of selection of the bacteria comes in function of a specific phenotype of interest, allows for positive functional selection of the variants	Large populations of mutator stains are difficult to stabilize after a prolonged selection pressure, as they tend to accumulate several spontaneous mutations. Introducing the mutator plasmid to a regular stain prevents this	Prepare mutator plasmid with gene of interest Transform competent cells with mutator plasmid Cultivate cells by dilutions and calculate the mutation frequency with the number of resistant colonies divided by the total of plated cells Grow cells with the mutator plasmid

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
SOE (for site-directed mutagenesis or site-saturation mutagenesis)	An asexual methodology for the generation of specific or random mutations of a codon at one targeted position	The PCRs with forward and mutagenic primers, and with reverse and mutagenic primers are carried out separately. Then, the PCR fragment products that overlap at the mutated codon at the specific sequence are stitched together by another PCR	Helps to specifically map the active site and study structure–function relationships. In several examples, when performed on key residues it has allowed optimization of the active site	The method requires the presence of restriction sites available at both ends of the target DNA sequence. If not present, silent mutations should be performed to introduce them Variations of this method involve whole plasmid amplification that requires a pair of complementary primers encoding the directed mutation or the degenerated codon. The PCR product can be directly transformed into competent cells. Thus, simplifying the process intro a single step. However, Dpn1 digestion of WT template must be carried out and occasionally, primer dimer formation can occur	Primers are designed to anneal to upstream and downstream of the target codon sequence Forward and reverse primers with an NNN/NNG codon are designed to anneal at the corresponding target codon of the gene Two PCRs are run separately to amplify the segments with overlapping target codon Run a PCR with both products of the previous PCR with upstream and downstream primers Purify and ligate PCR product into vector Transformation of cells
Multiple-site saturation mutagenesis	An asexual method for the generation of random mutations at various predetermined target positions	Use of long mutagenic primers containing NNN/NNG at every mutated site	It is an extension of SOE, thus, the purified plasmid obtained with a single-site mutation can used to as a template generate a double-site mutation in a plasmid	Limited to a few neighbouring codons per primer, since experimentally the efficiency decreases for each additional	Design primers and follow same steps as in SOE or plasmid amplification Purify amplified the single-site DNA fragments or plasmids Use plasmids as the template for the next rounds to induce more sites

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
RID	An asexual method for random mutations, produces insertions and deletions in a random manner along the entire gene	A circular dsDNA gene fragment is edited with linker and digested to produce circular ssDNA, which is randomly cleaved at single positions by the Ce(IV)-EDTA complex. Then, the fragments are ligated on their end with anchors (which may include the sequence to be inserted), and finally the anchors are digested leaving the gene fragments with the insertion	Allows the deletion of ≤ 16 bases of any site of the target gene, following insertion of a random or predetermined sequence of any number of bases at the same position	Variations in concentrations of various reagents and incubation times will cause unexpected deletions or inter-molecular ligation Ce(IV)-EDTA complex tends to precipitate after standing, producing loss of DNA cleavage activity. Use fresh solution	Circular dsDNA is cleaved, and a linker is inserted on the top strand of DNA Cyclization of gene fragment followed by degradation of unlinked antisense DNA Obtaining of ssDNA and treatment with Ce(IV)-EDTA Ligation of fragments of ssDNA to 5′ and 3′-anchors containing sequence for insertion at the 5′-end Amplification by regular PCR Digestion of anchors leaving the inserted sequence Cyclization of dsDNA Digestion of circular dsDNA with specific enzymes to create linear dsDNA with cloning sites for vector insertion Transformation of cells
Targeted random oligonucleotide mutagenesis	It generates random mutations into a specific region of a gene	Mutations are encoded on an artificial oligonucleotide that is partially degenerated, but it is still specific to a sequence in the gene, allowing for exploration of mutations within a specific area	Rapid assembly of libraries Allows for positive selection of variants Can be applied to prokaryotic and eukaryotic systems to produce assemblies of mutant sequences at a specific site, to explore its functionality and select for novel phenotypes not observed in nature	The degree of randomization depends on the extension and conservation of the region It is not possible to explore all combination of nucleotides even in a small region due to the vast number of probable sequences. However, enzyme activity is usually hindered with excessive mutations, thus, the method is convenient	Creation of a library of constructs with randomized dsDNA oligonucleotides and with restriction sites for later ligation into the gene of interest Preparation of the vector with the gene of interest by elimination of WT sequence of interest through a stuffer fragment with the same restriction sites as the randomized oligonucleotides Digest the stuffer fragment, purify the vector and ligate with randomized oligonucleotides Transformation of cells with vector

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
DNA shuffling	A sexual in vitro homologous recombination method, it searches for the creation of chimaeras by in vitro recombination of homologous genes	The homologous genes are fragmented and then reassembled by a PCR, and then ligated into a vector	Allows a recombination of larger sequences of a protein	There is a high probability of sequence uneven coverage at gene reassembly Fragment size by controlling digestion time and selection of size in agarose gel is important for reassembly In the amplification step, the number of cycles and the amount of template are critical and should be standardized Not fully random as DNase I usually cleaves at sites next to a pyrimidine nucleotide	Parental genes are digested with DNase I The digestion fragments are purified and recombined with a primer-less PCR Subsequent amplification by regular PCR with inner and outer primers generates full-length chimaeras for cloning into a vector Transformation of cells
Family shuffling with ssDNA	A sexual in vitro homologous recombination method, it produces homologous (heteroduplex) chimeric genes	Involves the recombination of homologous gene fragments that are produced by the restriction enzyme digestion of ssDNA	Allows assessment of a functionally larger sequence of a protein and overcomes drawbacks of dsDNA shuffling due to the annealing of fragments derived from the same parental genes cDNA or genomic DNA may be used	In many cases the lambda exonuclease cannot digest completely dsDNA to ssDNA. It may also degrade ssDNA, specially at high concentration and long-time of reaction It is recommended to purify the ssDNA with gel electrophoresis DNA polymerases with proofreading activity can minimize introduction of point mutations Due to minor binding affinity of ethidium bromide to ssDNA, thus, when calculating amount of DNA by ethidium bromide intensity, it will always be lower than the real amount Primers should include restriction sites for directional sub-cloning	Genes of dsDNA with one-strand 5′-end phosphorylated are amplified by PCR (the 5′ ends between genes are complementary to each other) Gene's ssDNA is obtained with lambda exonuclease digestion ssDNA of the set of homologous genes are mixed at equal proportions and digested with DNase I Gel electrophoresis is performed to select ssDNA fragment sizes The ssDNA pieces are reassembled Amplification by complementary primers with restriction sites, but only one primer is 5′-phosphoritated Cloning of DNA into a vector Transformation of cells

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
RPR	A sexual in vitro homologous recombination method, it is designed to produce homologous chimeric genes by random hexanucleotide-primers	Based on the stitching of DNA fragments produced by random hexanucleotide-primers. The fragments are reassembled with cycles of denaturation, annealing and enzyme polymerization to obtain the full gene sequence	RPR may use dsDNA, ssDNA or RNA templates and requires less parental DNA than DNA shuffling The length of DNA is not of concern. Also allows further mutations by misincorporation and mispriming. Overcomes sequence bias and random cleavage of DNA shuffling	Limited by the DNA composition In specific cases, a DNA polymerase with 3′ to 5′ exonuclease activity should be used in random-priming synthesis to avoid faulty incorporation of nucleotides	PCR with random hexamers Purification of DNA fragments Reassembly of fragments by a thermocycling process Amplification of chimerical genes by regular PCR Cloning of genes into a vector Transformation of bacteria
StEP	A sexual in vitro homologous recombination method, it implies to recombine DNA sequences using below-standard settings in vitro to produce full-length chimeric genes	Uses cross hybridization during partial amplification of DNA due to non-ideal conditions of extension time and temperature. Then, after digestion of DNA template, a second PCR amplifies the chimeric products	In comparison with DNA shuffling, it is simpler and does not require significant effort Uses only flanking primers and no nuclease is required	Limited by DNA template composition Too short or large plasmids and DNA templates should be avoided If the desirable range of temperature is not maintained during PCR process, it will produce non-specific annealing, producing undesirable products	PCR with primers, brief extension time and suboptimal temperature Template Dpn1 digestion Amplification of hybrid products by conventional PCR Cloning of hybrid genes into a vector Transformation of cells
RACHITT	A sexual in vitro homologous recombination method, its objective is to obtain chimeric genes by crossovers through a single round of gene shuffling	A bottom strand transient template containing uracil is hybridized with top strand DNA donor fragments that are stabilized by a single long annealing step. The unannealed and flap termini of donor fragments are cleaved by nucleases. The remaining fragments are ligated and, after treatment with UDG (to prevent amplification of DNA template) the hybrid gene is amplified	Multiple crossovers occur at high frequency and ssDNA can be used as template to ensure crossovers	It is a labour-intensive method The parental gene has a significant bias in the incorporation of the donor fragments The ratio of concentrations of donor DNA fragments will alter the final library An anchor oligonucleotide is used to protect the template's 5′ termini from nucleases Not fully random as DNase I usually cleaves at sites next to a pyrimidine nucleotide	Production of the bottom strand transient template and top strand DNA donor fragments Hybridization of fragments with template at a relatively high ionic strength Digestion of 5′ and 3′ flaps with DNase I nuclease and ligation of fragments to obtain a continuous chimeric top strand Digestion of template by UDG Amplification of ssDNA Ligation into a vector Transformation of bacteria

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
ITCHY library of two genes	A sexual in vitro non-homologous recombination method. Involves exon shuffling and it intends to generate a library of hybrid genes with internal deletions and duplications by sequence truncation-ligating methods	Creation of a library of hybrid genes by base truncation of their WT sequence, either by a time-dependent Exo III digestion or by addition of α-phosphorothioates to the sequence (THIOITCHY)	Allows creation of a comprehensive fusion library between fragments of genes without any sequence dependency, in order to create proteins with improved or novel properties	Variants contain only two genes and one crossover Time-dependent truncation requires multiple time-dependent sampling, while α-phosphorothioates is effortless and may combine random mutagenesis In theory, time-dependent digestion has a higher control of truncation and of parental length fusion, but this only occurs under optimal experimental conditions Use of pSALect vector for in-frame selection of DNA sequences	**Time-dependent Exo III digestion** Two genes cloned in tandem are linearized and digested by Exo III from their 5′ and 3′ ends in a time-dependent fashion Mung bean nuclease digestion The genes are uncoupled, and the linear DNA is ligated as a single gene with its vector Transformation of cells **α-phosphorothioates** Two genes cloned in tandem gene are amplified by PCR using a small amount of αS-dNTPs Exo III digestion, followed by mung bean nuclease digestion are performed, which is prevented by the random αS-dNMPs inserted in the DNA sequence Re-ligation of 5′ and 3′ ends Transformation of cells
SCRATCHY	A sexual in vitro non-homologous recombination method Involves exon shuffling and it aims to combine libraries of hybrid genes from more than one parental DNA sequence	Combines ITCHY and DNA shuffling techniques	Overcomes problems of large sequence deletions and insertions, as well as disruption of the correct nucleotide-codon reading frame for the fusion of genes	The method is labour-intensive Careful optimization of the conditions such as time, temperature, amount of nuclease is needed for optimal results It is important to reintroduce a STOP codon during the PCR either before DNA shuffling or after recombination of segments during amplification	Independent linearized vectors from the ITCHY libraries are separated by gel electrophoresis The desired fragment lengths are extracted from the gel, mixed together and cloned into a pSALect vector for in-frame selection The plasmid from selected colonies is obtained and used for DNA shuffling After digestion, recombination, amplification and ligation into a vector, Follows cell transformation

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
CLERY (and other techniques which recombine DNA in yeast)	A sexual in vivo homologous recombination method, it is designed to produce mosaic genes of eukaryotic hosts	Uses artificial techniques to introduce mutations, deletions, transitions, insertions and many others in a gene of interest; this is followed by in vivo recombination of homologous genes	Feasible for eukaryotic genes by combining in vitro and in vivo methods for DNA-recombination Increases recombination by using yeast capability of shuffling parental genes with just a few bases overlapping ($\geq$ 20 pb) and as many 200 nucleotides of homologous sequence	For some proteins, their initial secretion levels are not high enough to perform directed evolution, but with some mutational strategies this problem can be solved Hyper-glycosylation in *S. cerevisiae* is abundant, in some cases it can be over 50%, which complicates protein purification and characterization; but by reducing their transit-time at the Golgi apparatus	Different artificial methods might be used to create a combinatorial gene library as DNA template Amplification of the combinatorial DNA template Preparation of DNA and suitable vector with overlapping ends In vivo cloning and recombining by yeast
NExT	Considered a sexual in vitro homologous recombination method, it involves obtaining of a gene library from cleaving and shuffling uridine-sites substituted in the original gene	Uridine is inserted in the gene sequence by a PCR and dUTP. This nucleotide allows cleavage of DNA fragments by enzymatic digestion and further chemical cleavage with a base. The fragments are re-ligated as in DNA shuffling	A software helps to calculate experimental conditions The technique allows for gene recombination with short length fragments Allows for low error rate and no contamination with unshuffled clones	AT-rich regions will produce several small fragments, and GC-rich regions will not present cleavages	PCR reaction with dUTP is performed Uracil-PCR product is gel purified Uracil is cleaved by UDG digestion Cleaving reaction with piperidine Purification of fragments and reassembly of these at increasing annealing temperatures Amplification of products by regular PCR Cloning into a vector and cell transformation
SHIPREC	A sexual in vitro non-homologous recombination method Involves exon shuffling and it searches for the creation of single-crossover hybrids genes of low homology between them, by using an incremental truncation method	The genes with no sequence homology cloned in tandem joined by a linker region with restriction sites. The dimer is digested to a particular size and ligated. Afterwards, the circular DNA can be directionally cloned into a vector by the restriction sites located in the linker region	Allows for recombining families of proteins with similar functions but different sequences, nearly < 70–80% different The fraction of bases in different genes are fused toc codify one functional protein maximized in length and folding	These chimaeras contain only two genes and one crossover Use of pSALect vector for in-frame selection of DNA sequences	Fusion of genes by a linker sequence with several restriction sites Digestion of construct by DNase I and production of blunt ends by S1 nuclease Ligation of blunt ends Digestion of linker restriction sites by restriction enzymes Cloning of hybrid gene, cell transformation

Table 3 (continued)

Technique	Objective	How does it work?	Advantages	Important notes	Basic steps
RETT	A sexual in vitro non-homologous recombination method. Involves exon shuffling and uses either serial digestion with exonuclease or unidirectional extension of DNA by template switching to produce segments and generate hybrid genes	Various mRNAs are reverse-transcribed into ssDNAs. These work as templates for the segments to be produced for later assembling of the chimeric gene. If using serial deletion, the ssDNA is sequentially cleaved. If using combinatorial synthesis, the ssDNA is amplified by gene specific primers	The method does not use endonucleases to produce DNA segments It does not require suboptimal annealing and extension conditions	Use of pSALect vector for in-frame selection of DNA sequences It is important to reintroduce restriction sites and a STOP codon after recombination of segments during amplification of gene	RT-PCR of mRNA to produce ssDNA Production of fragments by reverse by serial deletion or unidirectional extension Gene fragments are then annealed to specific primer and are extended by PCR for a cycle Next cycle the ssDNA fragments anneal to another template and are extended The template switching, and annealing continues for all cycles until full-length hybrid ssDNA genes are obtained Cloning, and transformation

αS-dNMPs, deoxynucleoside 5′-(α-P-thio)monophosphate; αS-dNTPs, deoxynucleoside 5′-(α-P-thio)triphosphate; CLERY, Combinatorial Libraries Enhanced by Recombination in Yeast; dsDNA, double stranded DNA; ITCHY, Incremental Truncation for the Creation of Hybrid enzYmes; NExT, Nucleotide Exchange and Excision Technology DNA shuffling; RACHIT, RAndom CHImeragenesis on Transient Templates; RETT, Recombined Extension on Truncated Templates; RPR, Random-Priming in vitro Recombination; SHIPREC, Sequence Homology-Independent Protein RECombination; RID, Random Insertions and Deletions; SOE, Sequence Overlap Extension; ssDNA, single stranded DNA; StEP, Staggered Extension Process; UDG, uracil-DNA glycosilase) [88, 91–99]

Table 4 Instrumental techniques to study proteins

Technique	Description	Refs.
CD	It studies the changes in secondary structure of proteins with changes in the environmental conditions or during interaction with other molecules, by observing differences in the absorption of both left-handed and right-handed circularly polarized light in chiral centres of the molecule. In proteins, different structural elements have characteristic CD spectra	[106]
Pulse proteolysis	It assesses protein thermodynamic stability by comparing the proteolysis rates of unfolded proteins in equilibrium mixtures of folded and unfolded proteins. The proteolytic susceptibility of a protein depends on changes in its conformation, thus, by adding a protease to a purified protein sample and assessing the amount of remaining amount of folded protein	[107]
ITC	It assesses thermodynamical changes in a reaction such as affinity, enthalpy, entropy, free energy and stoichiometry by measuring the heat released when titrating one reactant with another under isothermal conditions. The parameters help to determine bonding, structure and rearrangements as a function of thermodynamics	[108]
UV–vis and fluorescence spectroscopy	These spectrophotometric techniques follow the course of an enzyme reaction by measuring the changes in the intensity of the light absorbed or emitted by the reaction solution. These rely on the property of a molecule to absorb UV–visible light which can either absorb it (transition of electrons from the ground state to the excited state) or emit light at a lower wavelength (transition of electrons from the excited state to the ground state). These are useful not only for quantification, but also to observe conformational changes by alterations in aromatic residues microenvironments, and analyse protein stability, it is based on the observation that the second derivative UV spectrum of aromatic residues shows changes according to their interactions with the local environment and solvent accessibility	[109]
HPLC	It separates, detects and quantifies components in a liquid mixture by pumping the mixture dissolved in the mobile phase through a stationary phase (column) that given its physicochemical properties, interacts differently with each compound of the mixture and results in their separation. Detection is carried out by a suitable method (electrochemistry, UV–vis, refractive index, diode array, etc)	[110]
Viscosity	Measured by a viscometer. The viscosity and density of the medium can influence the viscosity-dependent structural changes by friction against the protein, and thus, the rate-limiting step of an enzymic reaction (related to the proportion of molecular transitions over the energy barrier prior to product generation)	[111]
MS	It measures the mass-to-charge ratio of a sample by ionizing it, then the ions are sorted either by their acceleration by an electric or by their deflection by a magnetic field and are detected as a function of their size to charge ratio. It is a fast and sensitive technique that does not require chromophores or isotopic labelling. Moreover, allows the detection of proteolytic fragments and fragmentation of molecular ions, that allows identification of molecules, proteins and their structural changes. When coupled to HPLC it is a powerful detector	[112]
NMR	It is a non-destructive technique that provides structural information and dynamical changes in a molecule by detection of changes in the resonant frequency of spin ½ active nuclei in the molecule when placed under the influence of an external magnetic field. The absorbed energy to is a quantized process and must be the same as the difference in energy between the resonant frequency of both states involved. Thus, the variables involved are the magnetic field strength (T) and the frequency of radiation for resonance (Hz) Thanks to novel TROSY methods and labelling techniques, studies of large molecules or complexes are possible	[113]
DLS	It is measures the size, shape, and other structural characteristics of particles within the sub-micron region, and the diffusion of molecules in a solution through measurement of fluctuations in light intensity by light scattering caused by the Brownian Motion of particles upon radiation with a monochromatic wave of light. In enzymatic reactions, it assesses the effects of various environmental conditions (ion concentration, pH, and temperature) on the structure, association, aggregation, micellization, and activity of the enzyme	[114, 115]

CD, circular dichroism; DLS, dynamic light scattering; HPLC, high performance liquid chromatography; ITC, isothermal titration calorimetry; MS, mass spectrometry; NMR, nuclear magnetic resonance; TROSY, transverse relaxation optimized spectroscopy)

linear relationship between protein solubility and the addition of another co-solute of low viscosity and insignificant denaturation effect, this way the amount of precipitated protein increases in proportion to the amount of polymer [121, 123].

Modification of specificity vs selectivity of the enzyme

Enzymes can be selective catalysts and bind preferably to their substrate, but at the same time, they may be able to bind with other compounds with less affinity. On the other hand, the specificity of an enzyme arises from the exclusive reaction with a definite substrate. It is proposed that a good substrate can cause a conformational change in the active site of the enzyme, and similarly, the enzyme can induce a transition state in the substrate so that both can react; consequently, the specificity aligns with the rate in which a particular substrate reacts with the enzyme rather than the tendency between the enzyme and an analyte to bind each other [124].

To measure the rate of reaction with a specific substrate and under the assumption that enzymes operate under steady-state conditions, k_{cat} and K_M are experimentally calculated, and the ratio k_{cat}/K_M is often used to assess enzymatic specificity and to compare the relative rates of reaction of the given substrates [125–127]. A comprehensive treatise on k_{cat} and K_M is beyond the scope of this review and for better explanation of these concepts suggested reviews are Cornish-Bowden [127] and Schnell [125].

If k_{cat}/K_M large (by large value of k_{cat} or small value of K_M) indicates optimum kinetics of the system and the high enzyme specificity and which includes the control for isomer specificity of substrate and product. One example is pyridoxal phosphate (PLP), a generalized cofactor that can catalyse multiple reactions like transamination, racemization and decarboxylation; in contrast, PLP-dependent enzymes present characteristic and unique features in their structure to be limited a specific reaction, which permits specific stereoelectronic effects and chemical conditions that are exploited by the catalytic mechanism of the enzyme [128–130].

Another case to illustrate enzyme specificity is histone modification, where these reactions for post-translational modifications not only are specific, but also reversible, dynamic and highly controlled as they intervene in the interconversion of euchromatin to heterochromatin and modify gene expression (in eukaryotes) [131, 132]. Methyltransferases (MTases) are part of these genome-editing tools and use cofactor *S*-adenosyl-L-methionine (SAM) to mono-, di- or tri-methylate Arg and Lys residues of histones [132–134]. A conserved SET domain is shared by most types of Lys-MTases, this domain has a hydrophobic structure known as the Lys-channel, where the Lys reacts with the cofactor and the nucleophilic attack to SAM for methylation occurs. This structure also confers the specificity of the MTase to mono-, or di- or tri-methylate the terminal amino group of Lys. Structural analysis of the individual residues that define the internal size of the Lys-channel, observes that specific amino acids at a conserved structural position, allow specificity of the methylated states of the products of Lys-MTases. This conserved position is denominated the Tyr/Phe switch, as either amino acid can be found at this position and determine if the enzyme is able to transfer to its substrate up to three methyls by having a Phe, or if the enzyme is limited to monomethylation by having a Tyr. Site-directed mutagenesis in the Lys-channel to exchange Phe to Tyr or the opposite allows either including or excluding various degrees of methylation in the Lys without altering the general catalysis of the MTase [135–147].

Furthermore, though the substrate binding sequence of some MTases is present in various hundreds of non-histone cellular proteins, in vivo studies indicate that the substrate amino acid sequence alone is not enough for activity, and that the adoption of a precise conformation of the protein structure is required; consequently, just a few non-histone proteins are substrates of MTases [148–150].

It is possible to modify and re-design the active site of an enzyme to improve or even modify its substrate specificity and enantiomeric selectivity. Since enantiomers are mirror images, mutations must be designed as if the enzyme has to recognize and bind a whole new molecule [151, 152]. Physicochemical properties, such as high hydrostatic pressure, have been used to change stereoselectivity of an enzymatic reaction with no need of mutagenesis [153, 154]. Rational design in protein engineering has been used to change specificity and additionally allows understanding of the molecular basis of substrate specificity and chiral selectivity of enzymes [155, 156]. Mutagenesis through directed evolution is a viable option to change specificity as long as the experiment design and selection method are adequate, even if there is not much understanding of the structural features of the enzyme [91, 157].

Protein engineering has been applied to obtain orthogonality of enzymatic systems. Bioorthogonality is an important feature desired for artificial systems that urge to avoid any interactions with the inherent biochemical pathways in the cell. In orthogonal enzymatic labelling it is desired to selectively attach tags to proteins using an enzyme with higher affinity for a synthetic analogue cofactor than for the WT cofactor that might be highly abundant in the native cellular environment. This involves adjusting the structures in the enzyme-cofactor-substrate complex that permit the natural specificity, in order to re-design crucial interactions and thus enable the desired orthogonality [158–160]. Thus, the engineered enzyme in these bioorthogonal systems shows exclusive activity with an artificial substrate and cofactor, while it does not bind nor reacts towards the native molecules.

Conclusions and future perspectives

In this review, we have shown that protein engineering is a wide and developing area with an open range of methods and techniques that are constantly improving to yield the desired features in biocatalysts. This area combines both rational and empiric knowledge, along with various fields and analytical tools, and therefore lies at the interface between chemistry, biology and biotechnology. Moreover, advances in protein engineering and chemical proteomics not only propose applications to improve various areas, like environmental bioremediation or obtaining of new drugs, but also have the potential to improve our understanding of enzymes in aspects such as structure and activity, by providing both qualitative and quantitative information.

The improved knowledge in protein modification and biocatalysis enables research through the combination of computational and experimental methods, making protein engineering faster, cheaper and in various occasions with remarkable results; such as improved activity, stability, selectivity or even orthogonality beyond a traditional design. As aforementioned, it is a trend that bioinformatics and cheminformatics will continue to crossover and assist in the development of enzyme engineering.

With the novel goals in synthetic biology and biocatalysis, such as optimization of biosynthetic pathways, genetic circuits and genome engineering; novel tools and new knowledge will come forward to satisfy the needs of research and our understanding of life from a biological multilevel perspective that is the genome, to the transcriptome, to the proteome and epigenetics.

Acknowledgements We thank Jason Micklefield, Binuraj R. K. Menon, Anna-Winona Struck, Matthew Bennett, Mark Thomson and Brian Law for their support on the research work by Ph.D. Jimenez-Rosales at the University of Manchester. We also gratefully acknowledge financial support from both CONACYT scholarship 310271 and to SEP-PRODEP 511-6/17-9932.

Funding This work was supported by CONACYT doctoral funding 310271 and SEP-PRODEP award 511-6/17-9932 postdoctoral funding 993201.

Compliance with Ethical Standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Bugg, T. D. (2001). Enzymes: General properties. In John Wiley & Sons, Ltd (Ed.), *Encyclopedia of Life Sciences*. Chichester: Wiley. https://doi.org/10.1038/npg.els.0000709.
2. Baier, F., Copp, J. N., & Tokuriki, N. (2016). Evolution of enzyme superfamilies: Comprehensive exploration of sequence–function relationships. *Biochemistry, 55*(46), 6375–6388. https://doi.org/10.1021/acs.biochem.6b00723.
3. Pabis, A., Risso, V. A., Sanchez-Ruiz, J. M., & Kamerlin, S. C. (2017). Cooperativity and flexibility in enzyme evolution. *Current Opinion in Structural Biology, 48*, 83–92. https://doi.org/10.1016/j.sbi.2017.10.020.
4. Cobb, R. E., Chao, R., & Zhao, H. (2013). Directed evolution: past, present and future. *AIChE Journal American Institute of Chemical Engineers, 59*(5), 1432–1440. https://doi.org/10.1002/aic.13995.
5. Lane, M. D., & Seelig, B. (2014). Advances in the directed evolution of proteins. *Current Opinion in Chemical Biology, 22*, 129–136. https://doi.org/10.1016/j.cbpa.2014.09.013.
6. Magliery, T. J. (2015). Protein stability: computation, sequence statistics, and new experimental methods. *Current Opinion in Structural Biology, 33*, 161–168. https://doi.org/10.1016/j.sbi.2015.09.002.
7. Kulshreshtha, S., Chaudhary, V., Goswami, G. K., & Mathur, N. (2016). Computational approaches for predicting mutant protein stability. *Journal of Computer-Aided Molecular Design, 30*(5), 401–412. https://doi.org/10.1007/s10822-016-9914-3.
8. Baweja, M., Nain, L., Kawarabayasi, Y., & Shukla, P. (2016). Current technological improvements in enzymes toward their biotechnological applications. *Frontiers in Microbiology, 7*, 965. https://doi.org/10.3389/fmicb.2016.00965.
9. Anbar, M., Gul, O., Lamed, R., Sezerman, U. O., & Bayer, E. A. (2012). Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. *Applied and Environmental Microbiology, 78*(9), 3458–3464. https://doi.org/10.1128/AEM.07985-11.
10. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., & Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. *Science, 311*(5760), 484–489. https://doi.org/10.1126/science.1114736.
11. Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nature Methods, 11*(5), 508–520. https://doi.org/10.1038/nmeth.2926.
12. Zheng, X., Xing, X.-H., & Zhang, C. (2017). Targeted mutagenesis: A sniper-like diversity generator in microbial engineering. *Synthetic and Systems Biotechnology, 2*(2), 75–86. https://doi.org/10.1016/j.synbio.2017.07.001.
13. Kingsley, L. J., & Lill, M. A. (2015). Substrate tunnels in enzymes: structure–function relationships and computational methodology. *Proteins, 83*(4), 599–611. https://doi.org/10.1002/prot.24772.
14. Kuo, Y.-M., Henry, R. A., & Andrews, A. J. (2016). Measuring specificity in multi-substrate/product systems as a simple tool to investigate selectivity in vivo. *Biochimica et Biophysica Acta, 1864*(1), 70–76. https://doi.org/10.1016/j.bbapap.2015.08.011.
15. Samanta, S., & Mukherjee, S. (2017). Co-operative intra-protein structural response due to protein–protein complexation revealed through thermodynamic quantification: study of MDM2-p53 binding. *Journal of Computer-Aided Molecular Design, 31*(10), 891–903. https://doi.org/10.1007/s10822-017-0057-y.
16. Unterlass, J. E., Wood, R. J., Baslé, A., Tucker, J., Cano, C., Noble, M. M. E., & Curtin, N. J. (2017). Structural insights into the enzymatic activity and potential substrate promiscuity of human 3-phosphoglycerate dehydrogenase (PHGDH). *Oncotarget, 8*(61), 104478–104491. https://doi.org/10.18632/oncotarget.22327.
17. Jala, V. R., Appaji Rao, N., & Savithri, H. S. (2003). Identification of amino acid residues, essential for maintaining the tetrameric structure of sheep liver cytosolic serine hydroxymethyltransferase, by targeted mutagenesis. *Biochemical Journal, 369*(Pt 3), 469–476. https://doi.org/10.1042/BJ20021160.

18. Chatterjee, R., & Yuan, L. (2006). Directed evolution of metabolic pathways. *Trends in Biotechnology, 24*(1), 28–38. https://doi.org/10.1016/j.tibtech.2005.11.002.
19. Socha, R. D., & Tokuriki, N. (2013). Modulating protein stability—directed evolution strategies for improved protein function. *FEBS Journal, 280*(22), 5582–5595. https://doi.org/10.1111/febs.12354.
20. Taylor, J. L., Price, J. E., & Toney, M. D. (2015). Directed evolution of the substrate specificity of dialkylglycine decarboxylase. *Biochimica et Biophysica Acta, 1854*(2), 146–155. https://doi.org/10.1016/j.bbapap.2014.12.003.
21. Ratananikom, K., Choengpanya, K., Tongtubtim, N., Charoenrat, T., Withers, S. G., & Kongsaeree, P. T. (2013). Mutational analysis in the glycone binding pocket of *Dalbergia cochinchinensis* β-glucosidase to increase catalytic efficiency toward mannosides. *Carbohydrate Research, 373*, 35–41. https://doi.org/10.1016/j.carres.2012.10.018.
22. Suarez, S. C., Beardslee, R. A., Toffton, S. M., & McCulloch, S. D. (2013). Biochemical analysis of active site mutations of human polymerase η. *Mutation Research, 745–746*, 46–54. https://doi.org/10.1016/j.mrfmmm.2013.03.001.
23. Khan, S., Pozzo, T., Megyeri, M., Lindahl, S., Sundin, A., Turner, C., & Karlsson, E. N. (2011). Aglycone specificity of *Thermotoga neapolitana* β-glucosidase 1A modified by mutagenesis, leading to increased catalytic efficiency in quercetin-3-glucoside hydrolysis. *BMC Biochemistry, 12*, 11. https://doi.org/10.1186/1471-2091-12-11.
24. Weng, M., Zheng, Z., Bao, W., Cai, Y., Yin, Y., & Zou, G. (2009). Enhancement of oxidative stability of the subtilisin nattokinase by site-directed mutagenesis expressed in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1794*(11), 1566–1572. https://doi.org/10.1016/j.bbapap.2009.07.007.
25. Kobayashi, J., Furukawa, M., Ohshiro, T., & Suzuki, H. (2015). Thermoadaptation-directed evolution of chloramphenicol acetyltransferase in an error-prone thermophile using improved procedures. *Applied Microbiology and Biotechnology, 99*(13), 5563–5572. https://doi.org/10.1007/s00253-015-6522-4.
26. Seo, J.-H., Kim, H.-H., Jeon, E.-Y., Song, Y.-H., Shin, C.-S., & Park, J.-B. (2016). Engineering of Baeyer–Villiger monooxygenase-based *Escherichia coli* biocatalyst for large scale biotransformation of ricinoleic acid into (Z)-11-(heptanoyloxy) undec-9-enoic acid. *Scientific Reports*. https://doi.org/10.1038/srep28223.
27. Capone, S., Ćorajević, L., Bonifert, G., Murth, P., Maresch, D., Altmann, F., Herwig, C., & Spadiut, O. (2015). Combining protein and strain engineering for the production of glyco-engineered horseradish peroxidase C1A in *Pichia pastoris*. *International Journal of Molecular Sciences, 16*(10), 23127–23142. https://doi.org/10.3390/ijms161023127.
28. Larue, K., Melgar, M., & Martin, V. J. J. (2016). Directed evolution of a fungal β-glucosidase in *Saccharomyces cerevisiae*. *Biotechnology for Biofuels, 9*. https://doi.org/10.1186/s13068-016-0470-9.
29. Lian, J., Li, Y., HamediRad, M., & Zhao, H. (2014). Directed evolution of a cellodextrin transporter for improved biofuel production under anaerobic conditions in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering, 111*(8), 1521–1531. https://doi.org/10.1002/bit.25214.
30. Glenn, W. S., Nims, E., & O'Connor, S. E. (2011). Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *Journal of the American Chemical Society, 133*(48), 19346–19349. https://doi.org/10.1021/ja2089348.
31. Jain, K., Warmack, R. A., Debler, E. W., Hadjikyriacou, A., Stavropoulos, P., & Clarke, S. G. (2016). Protein arginine methyltransferase product specificity is mediated by distinct active-site architectures. *The Journal of Biological Chemistry, 291*(35), 18299–18308. https://doi.org/10.1074/jbc.M116.740399.
32. Chiang, C.-H., Grauffel, C., Wu, L.-S., Kuo, P.-H., Doudeva, L. G., Lim, C., Shen, C. K., & Yuan, H. S. (2016). Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Scientific Reports, 6*, 21581. https://doi.org/10.1038/srep21581.
33. Cohen, I., Kayode, O., Hockla, A., Sankaran, B., Radisky, D. C., Radisky, E. S., & Papo, N. (2016). Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *The Biochemical Journal, 473*(10), 1329–1341. https://doi.org/10.1042/BJ20151410.
34. Clark, D. P., & Pazdernik, N. J. (2016). Chapter 11: Protein engineering. Boston: Academic Cell. (pp. 365–392)
35. Lv, B., Sun, H., Huang, S., Feng, X., Jiang, T., & Li, C. (2017). Structure-guided engineering of the substrate specificity of a fungal β-glucuronidase toward triterpenoid saponins. *The Journal of Biological Chemistry, 293*(2), 433–443. https://doi.org/10.1074/jbc.M117.801910.
36. He, R., Reyes, A. C., Amyes, T. L., & Richard, J. P. (2018). Enzyme architecture: the role of a flexible loop in activation of glycerol-3-phosphate dehydrogenase for catalysis of hydride transfer. *Biochemistry*. https://doi.org/10.1021/acs.biochem.7b01282.
37. Reyes, A. C., Amyes, T. L., & Richard, J. P. (2016). Enzyme architecture: a startling role for Asn270 in glycerol 3-phosphate dehydrogenase-catalyzed hydride transfer. *Biochemistry, 55*(10), 1429–1432. https://doi.org/10.1021/acs.biochem.6b00116.
38. Gromiha, M. M. (2010). Chapter 6—protein stability. In *Protein Bioinformatics* (pp. 209–245). Singapore: Academic Press. https://doi.org/10.1016/B978-8-1312-2297-3.50006-0.
39. Martínez Cuesta, S., Rahman, S. A., Furnham, N., & Thornton, J. M. (2015). The classification and evolution of enzyme function. *Biophysical Journal, 109*(6), 1082–1086. https://doi.org/10.1016/j.bpj.2015.04.020.
40. Tsien, R. Y. (1998). The green fluorescent protein. *Annual Review of Biochemistry, 67*(1), 509–544. https://doi.org/10.1146/annurev.biochem.67.1.509.
41. Heim, R., & Tsien, R. Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology: CB, 6*(2), 178–182.
42. 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., Devine, P. N., & Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science, 329*(5989), 305–309. https://doi.org/10.1126/science.1188934.
43. 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., Ravichandran, R., & Baker, D. (2017). Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature, 552*(7685), 415–420. https://doi.org/10.1038/nature25157.
44. Kwong, P. D., & Mascola, J. R. (2018). HIV-1 vaccines based on antibody identification, B cell ontogeny, and epitope structure. *Immunity, 48*(5), 855–871. https://doi.org/10.1016/j.immuni.2018.04.029.
45. Olivera-Nappa, A., Andrews, B. A., & Asenjo, J. A. (2011). Mutagenesis Objective Search and Selection Tool (MOSST): an algorithm to predict structure-function related mutations in proteins. *BMC Bioinformatics, 12*(1), 122. https://doi.org/10.1186/1471-2105-12-122.
46. Currin, A., Swainston, N., Day, P. J., & Kell, D. B. (2015). Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical*

Society Reviews, 44(5), 1172–1239. https://doi.org/10.1039/C4CS00351A.
47. Tiwari, V. (2016). In vitro engineering of novel bioactivity in the natural enzymes. *Frontiers in Chemistry, 4*(39). https://doi.org/10.3389/fchem.2016.00039.
48. Taylor, A. I., & Holliger, P. (2015). Directed evolution of artificial enzymes (XNAzymes) from diverse repertoires of synthetic genetic polymers. *Nature Protocols, 10*(10), 1625–1642. https://doi.org/10.1038/nprot.2015.104.
49. Dalby, P. A. (2011). Strategy and success for the directed evolution of enzymes. *Current Opinion in Structural Biology, 21*(4), 473–480. https://doi.org/10.1016/j.sbi.2011.05.003.
50. 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.
51. 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.
52. Acevedo-Rocha, C. G., Hoebenreich, S., & Reetz, M. T. (2014). Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods in Molecular Biology, 1179*, 103–128. https://doi.org/10.1007/978-1-4939-1053-3_7.
53. Parra, L. P., Agudo, R., & Reetz, M. T. (2013). Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *Chembiochem: A European Journal of Chemical Biology, 14*(17), 2301–2309. https://doi.org/10.1002/cbic.201300486.
54. Cai, Y., Bhuiya, M.-W., Shanklin, J., & Liu, C.-J. (2015). Engineering a monolignol 4-*O*-methyltransferase with high selectivity for the condensed lignin precursor coniferyl alcohol. *The Journal of Biological Chemistry, 290*(44), 26715–26724. https://doi.org/10.1074/jbc.M115.684217.
55. Chaparro-Riggers, J. F., Polizzi, K. M., & Bommarius, A. S. (2007). Better library design: data-driven protein engineering. *Biotechnology Journal, 2*(2), 180–191. https://doi.org/10.1002/biot.200600170.
56. What is Rosetta@home? (2018). *Rosetta@home*. Retrieved April 28, 2018 from https://boinc.bakerlab.org/rosetta/rah/rah_about.php.
57. Niu, R.-J., Zheng, Q.-C., Zhang, J.-L., & Zhang, H.-X. (2013). Molecular dynamics simulations studies and free energy analysis on inhibitors of MDM2-p53 interaction. *Journal of Molecular Graphics & Modelling, 46*, 132–139. https://doi.org/10.1016/j.jmgm.2013.10.005.
58. Chen, J., Wang, J., Zhang, Q., Chen, K., & Zhu, W. (2015). Probing origin of binding difference of inhibitors to MDM2 and MDMX by polarizable molecular dynamics simulation and QM/MM-GBSA calculation. *Scientific Reports, 5*, 17421. https://doi.org/10.1038/srep17421.
59. Radchenko, T., Brink, A., Siegrist, Y., Kochansky, C., Bateman, A., Fontaine, F., Morettoni, L., & Zamora, I. (2017). Software-aided approach to investigate peptide structure and metabolic susceptibility of amide bonds in peptide drugs based on high resolution mass spectrometry. *PLoS ONE, 12*(11), e0186461. https://doi.org/10.1371/journal.pone.0186461.
60. Zhang, L., Ai, H.-X., Li, S.-M., Qi, M.-Y., Zhao, J., Zhao, Q., & Liu, H.-S. (2017). Virtual screening approach to identifying influenza virus neuraminidase inhibitors using molecular docking combined with machine-learning-based scoring function. *Oncotarget, 8*(47), 83142–83154. https://doi.org/10.18632/oncotarget.20915.
61. Studer, R. A., Dessailly, B. H., & Orengo, C. A. (2013). Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochemical Journal, 449*(3), 581–594. https://doi.org/10.1042/BJ20121221.
62. Wójcikowski, M., Ballester, P. J., & Siedlecki, P. (2017). Performance of machine-learning scoring functions in structure-based virtual screening. *Scientific Reports, 7*, 46710. https://doi.org/10.1038/srep46710.
63. Alcalde, M. (2016). When directed evolution met ancestral enzyme resurrection. *Microbial Biotechnology, 10*(1), 22–24. https://doi.org/10.1111/1751-7915.12452.
64. Wheeler, L. C., Lim, S. A., Marqusee, S., & Harms, M. J. (2016). The thermostability and specificity of ancient proteins. *Current opinion in structural biology, 38*, 37–43. https://doi.org/10.1016/j.sbi.2016.05.015.
65. Gilis, D., & Rooman, M. (2000). PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Engineering, 13*(12), 849–856.
66. Dehouck, Y., Kwasigroch, J. M., Gilis, D., & Rooman, M. (2011). PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics, 12*, 151. https://doi.org/10.1186/1471-2105-12-151.
67. Cabrita, L. D., Gilis, D., Robertson, A. L., Dehouck, Y., Rooman, M., & Bottomley, S. P. (2007). Enhancing the stability and solubility of TEV protease using in silico design. *Protein Science, 16*(11), 2360–2367. https://doi.org/10.1110/ps.072822507.
68. Olivera-Nappa, A., Reyes, F., Andrews, B. A., & Asenjo, J. A. (2013). Cold adaptation, Ca2+ dependency and autolytic stability are related features in a highly active cold-adapted trypsin resistant to autoproteolysis engineered for biotechnological applications. *PLoS ONE*, 8(8). https://doi.org/10.1371/journal.pone.0072355.
69. 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 Sciences of the United States of America, 114*(13), E2624–E2633. https://doi.org/10.1073/pnas.1700269114.
70. Bedbrook, C. N., Yang, K. K., Rice, A. J., Gradinaru, V., & Arnold, F. H. (2017). Machine learning to design integral membrane channel rhodopsins for efficient eukaryotic expression and plasma membrane localization. *PLoS Computational Biology, 13*(10), e1005786. https://doi.org/10.1371/journal.pcbi.1005786.
71. Heinzelman, P., Snow, C. D., Smith, M. A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., & Arnold, F. H. (2009). SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *The Journal of Biological Chemistry, 284*(39), 26229–26233. https://doi.org/10.1074/jbc.C109.034058.
72. Pai, P. P., Ranjani, S. S. S., & Mondal, S. (2015). PINGU: PredIction of eNzyme catalytic residues usinG seqUence information. *PLoS ONE, 10*(8), e0135122. https://doi.org/10.1371/journal.pone.0135122.
73. Ko, J., Murga, L. F., André, P., Yang, H., Ondrechen, M. J., Williams, R. J., Agunwamba, A., & Budil, D. E. (2005). Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins, 59*(2), 183–195. https://doi.org/10.1002/prot.20418.
74. Tong, W., Williams, R. J., Wei, Y., Murga, L. F., Ko, J., & Ondrechen, M. J. (2008). Enhanced performance in prediction of protein active sites with THEMATICS and support vector machines. *Protein Science: A Publication of the Protein Society, 17*(2), 333–341. https://doi.org/10.1110/ps.073213608.
75. Folding@home Fighting disease with a world wide distributed super computer. (n.d.). Retrieved from https://foldingathome.org/.
76. Craveur, P., Joseph, A. P., Esque, J., Narwani, T. J., Noël, F., Shinada, N., Goguet, M., Sylvain, L., Poulain, P., & de

Brevern, A. G. (2015). Protein flexibility in the light of structural alphabets. *Frontiers in Molecular Biosciences*. https://doi.org/10.3389/fmolb.2015.00020.
77. Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology, 22*(11), 1399–1408. https://doi.org/10.1038/nbt1029.
78. Collinet, B., Hervé, M., Pecorari, F., Minard, P., Eder, O., & Desmadril, M. (2000). Functionally accepted insertions of proteins within protein domains. *Journal of Biological Chemistry, 275*(23), 17428–17433. https://doi.org/10.1074/jbc.M000666200.
79. Giri Rao, V. V. H., & Gosavi, S. (2018). On the folding of a structurally complex protein to its metastable active state. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1708173115.
80. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Isolating, cloning, and sequencing DNA. In *Molecular Biology of the Cell* (4th ed.). New York: Garland Science.
81. Celie, P. H., Parret, A. H., & Perrakis, A. (2016). Recombinant cloning strategies for protein expression. *Current Opinion in Structural Biology, 38*, 145–154. https://doi.org/10.1016/j.sbi.2016.06.010.
82. Gilis, D., McLennan, H. R., Dehouck, Y., Cabrita, L. D., Rooman, M., & Bottomley, S. P. (2003). In vitro and in silico design of α1-antitrypsin mutants with different conformational stabilities. *Journal of Molecular Biology, 325*(3), 581–589. https://doi.org/10.1016/S0022-2836(02)01221-4.
83. Addgene: Choosing a molecular cloning technique. (n.d.). Retrieved October 10, 2017, from https://www.addgene.org/plasmid-reference/cloning-choice/.
84. Applications|NEB. (n.d.). Retrieved October 13, 2017, from https://www.neb.com/applications.
85. Protein Expression and Purification Core Facility—Cloning—Cloning Methods—EMBL. (n.d.). Retrieved October 10, 2017, from https://www.embl.de/pepcore/pepcore_services/cloning/cloning_methods/index.html.
86. Gibson, D. G. (2009). Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research, 37*(20), 6984–6990. https://doi.org/10.1093/nar/gkp687.
87. Poluri, K. M., & Gulati, K. (2017). *Protein engineering techniques: Gateways to synthetic protein universe*. Springer, Singapore. Retrieved from http://www.springer.com/gp/book/9789811027314.
88. Arnold, F. H., & Georgiou, G. (Eds.)., (2003). *Directed evolution library creation: methods and protocols*. Humana Press, New York. Retrieved from http://www.springer.com/gp/book/9781588292858.
89. Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America, 91*(22), 10747–10751.
90. Moore, G. L., & Maranas, C. D. (2002). eCodonOpt: a systematic computational framework for optimizing codon usage in directed evolution experiments. *Nucleic Acids Research, 30*(11), 2407–2416.
91. Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics, 16*(7), 379. https://doi.org/10.1038/nrg3927.
92. Müller, K. M., Stebel, S. C., Knall, S., Zipf, G., Bernauer, H. S., & Arndt, K. M. (2005). Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Research, 33*(13), e117. https://doi.org/10.1093/nar/gni116.
93. Sieber, V., Martinez, 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.
94. Gonzalez-Perez, D., Garcia-Ruiz, E., & Alcalde, M. (2012). *Saccharomyces cerevisiae* in directed evolution. *Bioengineered Bugs, 3*(3), 172–177. https://doi.org/10.4161/bbug.19544.
95. Böttcher, D., & Bornscheuer, U. T. (2010). Protein engineering of microbial enzymes. *Current Opinion in Microbiology, 13*(3), 274–282. https://doi.org/10.1016/j.mib.2010.01.010.
96. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D., & Benkovic, S. J. (2001). Creating multiple-crossover DNA libraries independent of sequence identity. *Proceedings of the National Academy of Sciences, 98*(20), 11248–11253. https://doi.org/10.1073/pnas.201413698.
97. 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.
98. Tseng, W.-C., Lin, J.-W., Wei, T.-Y., & Fang, T.-Y. (2008). A novel megaprimed 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.
99. Lee, S. H., Ryu, E. J., Kang, M. J., Wang, E.-S., Piao, Z., Choi, Y. J., Jung, K. H., Jeon, J. Y., & Shin, Y. C. (2003). A new approach to directed gene evolution by recombined extension on truncated templates (RETT). *Journal of Molecular Catalysis B: Enzymatic, 26*(3), 119–129. https://doi.org/10.1016/j.molcatb.2003.05.001.
100. Frauenkron-Machedjou, V. J., Fulton, A., Zhu, L., Anker, C., Bocola, M., Jaeger, K.-E., & Schwaneberg, U. (2015). Towards understanding directed evolution: more than half of all amino acid positions contribute to ionic liquid resistance of *Bacillus subtilis* lipase A. *ChemBioChem, 16*(6), 937–945. https://doi.org/10.1002/cbic.201402682.
101. Johnston, C. A., Whitney, D. S., Volkman, B. F., Doe, C. Q., & Prehoda, K. E. (2011). Conversion of the enzyme guanylate kinase into a mitotic-spindle orienting protein by a single mutation that inhibits GMP-induced closing. *Proceedings of the National Academy of Sciences of the United States of America, 108*(44), E973–E978. https://doi.org/10.1073/pnas.1104365108.
102. Woods, D. F., & Bryant, P. J. (1993). ZO-1, DlgA and PSD-95/SAP90: homologous proteins in tight, septate and synaptic cell junctions. *Mechanisms of Development, 44*(2), 85–89. https://doi.org/10.1016/0925-4773(93)90059-7.
103. Wyganowski, K. T., Kaltenbach, M., & Tokuriki, N. (2013). GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *Journal of Molecular Biology, 425*(18), 3403–3414. https://doi.org/10.1016/j.jmb.2013.06.028.
104. Rehm, F. B. H., Chen, S., & Rehm, B. H. A. (2016). Enzyme engineering for in situ immobilization. *Molecules, 21*(10). https://doi.org/10.3390/molecules21101370.
105. Singh, R. K., Tiwari, M. K., Singh, R., & Lee, J.-K. (2013). From protein engineering to immobilization: promising strategies for the upgrade of industrial enzymes. *International Journal of Molecular Sciences, 14*(1), 1232–1277. https://doi.org/10.3390/ijms14011232.
106. Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols, 1*(6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.
107. Na, Y.-R., & Park, C. (2009). Investigating protein unfolding kinetics by pulse proteolysis. *Protein Science: A Publication of the Protein Society, 18*(2), 268–276. https://doi.org/10.1002/pro.29.
108. Dutta, A. K., Rösgen, J., & Rajarathnam, K. (2015). Using isothermal titration calorimetry to determine thermodynamic parameters of protein–glycosaminoglycan interactions. *Methods*

in Molecular Biology. (Clifton, N.J.), 1229, 315–324. https://doi.org/10.1007/978-1-4939-1714-3_25.
109. Penner, M. H. (2010). Ultraviolet, visible, and fluorescence spectroscopy. In *Food Analysis* (pp. 387–405). Boston: Springer. https://doi.org/10.1007/978-1-4419-1478-1_22.
110. Concepts and Principles of High Performance Liquid Chromatography. (2006). In *HPLC in Enzymatic Analysis* (pp. 13–40). Wiley, New York. https://doi.org/10.1002/9780470110591.ch2.
111. Uribe, S., & Sampedro, J. G. (2003). Measuring solution viscosity and its effect on enzyme activity. *Biological Procedures Online, 5*, 108–115. https://doi.org/10.1251/bpo52.
112. de Boer, A. R., Lingeman, H., Niessen, W. M. A., & Irth, H. (2007). Mass spectrometry-based biochemical assays for enzyme-inhibitor screening. *TrAC Trends in Analytical Chemistry, 26*(9), 867–883. https://doi.org/10.1016/j.trac.2007.08.004.
113. Lisi, G. P., & Loria, J. P. (2016). Solution NMR spectroscopy for the study of enzyme allostery. *Chemical Reviews, 116*(11), 6323–6369. https://doi.org/10.1021/acs.chemrev.5b00541.
114. Lorber, B., Fischer, F., Bailly, M., Roy, H., & Kern, D. (2012). Protein analysis by dynamic light scattering: methods and techniques for students. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology, 40*(6), 372–382. https://doi.org/10.1002/bmb.20644.
115. Gast, K., & Fiedler, C. (2012). Dynamic and static light scattering of intrinsically disordered proteins. *Methods in Molecular Biology (Clifton, N.J.), 896*, 137–161. https://doi.org/10.1007/978-1-4614-3704-8_9.
116. Kuznetsova, E., Proudfoot, M., Sanders, S. A., Reinking, J., Savchenko, A., Arrowsmith, C. H., Edwards, A. M., & Yakunin, A. F. (2005). Enzyme genomics: Application of general enzymatic screens to discover new enzymes. *FEMS Microbiology Reviews, 29*(2), 263–279. https://doi.org/10.1016/j.fmrre.2004.12.006.
117. Acker, M. G., & Auld, D. S. (2014). Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science, 1*(1), 56–73. https://doi.org/10.1016/j.pisc.2013.12.001.
118. Gasymov, O. K., Abduragimov, A. R., & Glasgow, B. J. (2014). Probing tertiary structure of proteins using single trp mutations with circular dichroism at low temperature. *The Journal of Physical Chemistry. B, 118*(4), 986–995. https://doi.org/10.1021/jp4120145.
119. Holthauzen, L. M. F., Auton, M., Sinev, M., & Rösgen, J. (2011). Protein stability in the presence of cosolutes. *Methods in Enzymology, 492*, 61–125. https://doi.org/10.1016/B978-0-12-381268-1.00015-X.
120. Monera, O. D., Kay, C. M., & Hodges, R. S. (1994). Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Science: A Publication of the Protein Society, 3*(11), 1984–1991.
121. Li Li, A. Kantor, & Nicholas Warne. (2013). Application of a PEG precipitation method for solubility screening: A tool for developing high protein concentration formulations. *Protein Science*, 22(8), 1118–1123.
122. Roodveldt, C., & Tawfik, D. S. (2005). Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *Protein Engineering Design and Selection, 18*(1), 51–58. https://doi.org/10.1093/protein/gzi005.
123. Polson, A., Potgieter, G. M., Largier, J. F., Mears, G. E. F., & Joubert, F. J. (1964). The fractionation of protein mixtures by linear polymers of high molecular weight. *Biochimica et Biophysica Acta (BBA)—General Subjects, 82*(3), 463–475. https://doi.org/10.1016/0304-4165(64)90438-6.
124. Klinman, J. P., Offenbacher, A. R., & Hu, S. (2017). Origins of enzyme catalysis: experimental findings for C–H activation, new models, and their relevance to prevailing theoretical constructs. *Journal of the American Chemical Society, 139*(51), 18409–18427. https://doi.org/10.1021/jacs.7b08418.
125. Schnell, S. (2014). Validity of the Michaelis–Menten equation–steady-state or reactant stationary assumption: that is the question. *The FEBS Journal, 281*(2), 464–472. https://doi.org/10.1111/febs.12564.
126. Xu, R., Gu, E., Zhou, Q., Yuan, L., Hu, X., Cai, J., & Hu, G. (2016). Effects of 22 novel CYP2D6 variants found in Chinese population on the metabolism of dapoxetine. *Drug Design, Development and Therapy, 10*, 687–696. https://doi.org/10.2147/DDDT.S97789.
127. Cornish-Bowden, A. (2013). The origins of enzyme kinetics. *FEBS Letters, 587*(17), 2725–2730. https://doi.org/10.1016/j.febslet.2013.06.009.
128. Eliot, A. C., & Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annual Review of Biochemistry, 73*(1), 383–415. https://doi.org/10.1146/annurev.biochem.73.011303.074021.
129. Catazaro, J., Caprez, A., Guru, A., Swanson, D., & Powers, R. (2014). Functional evolution of PLP-dependent enzymes based on active-site structural similarities. *Proteins, 82*(10), 2597–2608. https://doi.org/10.1002/prot.24624.
130. Yoshimura, T., Jhee, K. H., & Soda, K. (1996). Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes. *Bioscience, Biotechnology, and Biochemistry, 60*(2), 181–187. https://doi.org/10.1271/bbb.60.181.
131. Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., & Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes & Development, 24*(15), 1590–1595. https://doi.org/10.1101/gad.586710.
132. Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms, 1839*(12), 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007.
133. Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics, 33*, 245–254. https://doi.org/10.1038/ng1089.
134. Tessarz, P., & Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology, 15*(11), 703–708. https://doi.org/10.1038/nrm3890.
135. Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y., & Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *The Journal of Biological Chemistry, 280*(7), 5563–5570. https://doi.org/10.1074/jbc.M410483200.
136. Rathert, P., Zhang, X., Freund, C., Cheng, X., & Jeltsch, A. (2008). Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays. *Chemistry & Biology, 15*(1), 5–11. https://doi.org/10.1016/j.chembiol.2007.11.013.
137. Couture, J.-F., Collazo, E., Brunzelle, J. S., & Trievel, R. C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes & Development, 19*(12), 1455–1465. https://doi.org/10.1101/gad.1318405.
138. Xiao, B., Jing, C., Kelly, G., Walker, P. A., Muskett, F. W., Frenkiel, T. A., Martin, S. R., Sarma, K., Reinberg, D., Gamblin, S. J., & Wilson, J. R. (2005). Specificity and mechanism of

the histone methyltransferase Pr-Set7. *Genes & Development, 19*(12), 1444–1454. https://doi.org/10.1101/gad.1315905.
139. Wu, H., Siarheyeva, A., Zeng, H., Lam, R., Dong, A., Wu, X.-H., Li, Y., Schapira, M., Vedadi, M., & Min, J. (2013). Crystal structures of the human histone H4K20 methyltransferases SUV420H1 and SUV420H2. *FEBS Letters, 587*(23), 3859–3868. https://doi.org/10.1016/j.febslet.2013.10.020.
140. Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J., & Xiao, B. (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell, 111*(1), 105–115. https://doi.org/10.1016/S0092-8674(02)00964-9.
141. Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Molecular Cell, 12*(1), 177–185. https://doi.org/10.1016/S1097-2765(03)00224-7.
142. Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., & Cheng, X. (2002). Structure of the neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell, 111*(1), 117–127. https://doi.org/10.1016/S0092-8674(02)00999-6.
143. Trievel, R. C. (2004). Structure and function of histone methyltransferases. *Critical Reviews in Eukaryotic Gene Expression, 14*(3), 147–169. https://doi.org/10.1615/CritRevEukaryotGeneExpr.v14.i3.10.
144. Qian, C., & Zhou, M.-M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cellular and Molecular Life Sciences CMLS, 63*(23), 2755–2763. https://doi.org/10.1007/s00018-006-6274-5.
145. Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology, 6*(8), 227. https://doi.org/10.1186/gb-2005-6-8-227.
146. Kudithipudi, S., Dhayalan, A., Kebede, A. F., & Jeltsch, A. (2012). The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie, 94*(11), 2212–2218. https://doi.org/10.1016/j.biochi.2012.04.024.
147. Qian, C., Wang, X., Manzur, K., Sachchidanand, A. F., Zeng, L., & Zhou, M.-M. (2006). Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. *Journal of Molecular Biology, 359*(1), 86–96. https://doi.org/10.1016/j.jmb.2006.03.006.
148. Dhayalan, A., Kudithipudi, S., Rathert, P., & Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chemistry & Biology, 18*(1), 111–120. https://doi.org/10.1016/j.chembiol.2010.11.014.
149. Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., & Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Molecular Cell, 53*(2), 277–289. https://doi.org/10.1016/j.molcel.2013.12.005.
150. West, L. E., Roy, S., Lachmi-Weiner, K., Hayashi, R., Shi, X., Appella, E., Kutateladze, T. J., & Gozani, O. (2010). The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. *Journal of Biological Chemistry, 285*(48), 37725–37732. https://doi.org/10.1074/jbc.M110.139527.
151. Zhang, Y., Ye, J., & Liu, M. (2017). Enantioselective biotransformation of chiral persistent organic pollutants. *Current Protein & Peptide Science, 18*(1), 48–56.
152. Guo, F., Zhang, J., & Wang, C. (2017). Enantioselectivity in environmental safety and metabolism of typical chiral organic pollutants. *Current Protein & Peptide Science, 18*(1), 4–9.
153. Golub, M., Lehofer, B., Martinez, N., Ollivier, J., Kohlbrecher, J., Prassl, R., & Peters, J. (2017). High hydrostatic pressure specifically affects molecular dynamics and shape of low-density lipoprotein particles. *Scientific Reports, 7*, 46034. https://doi.org/10.1038/srep46034.
154. Makarov, A. A., Helmy, R., Joyce, L., Reibarkh, M., Maust, M., Ren, S., Mergelsberg, I., & Welch, C. J. (2016). Use of hydrostatic pressure for modulation of protein chemical modification and enzymatic selectivity. *Organic & Biomolecular Chemistry, 14*(19), 4448–4455. https://doi.org/10.1039/C6OB00550K.
155. Didier, R. (2000). *Understanding and Engineering the Enantioselectivity of Candida antarctica Lipase B towards sec-Alcohols* (Thesis). KTH Royal Institute of Technology, Stockholm.
156. Boersma, Y. L., Dröge, M. J., van der Sloot, A. M., Pijning, T., Cool, R. H., Dijkstra, B. W., & Quax, W. J. (2008). A novel genetic selection system for improved enantioselectivity of *Bacillus subtilis* lipase A. *ChemBioChem, 9*(7), 1110–1115. https://doi.org/10.1002/cbic.200700754.
157. Xiao, H., Bao, Z., & Zhao, H. (2015). High throughput screening and selection methods for directed enzyme evolution. *Industrial & Engineering Chemistry Research, 54*(16), 4011–4020. https://doi.org/10.1021/ie503060a.
158. Zhang, C., Kenski, D. M., Paulson, J. L., Bonshtien, A., Sessa, G., Cross, J. V., Templeton, D. J., & Shokat, K. M. (2005). A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. *Nature Methods, 2*(6), 435–441. https://doi.org/10.1038/nmeth764.
159. Jimenez-Rosales, A. (2015). *Methyltransferases as bioorthogonal labeling tools for proteins* (PhD thesis). The University of Manchester, Manchester, United Kingdom.
160. Allen, J. J., Lazerwith, S. E., & Shokat, K. M. (2005). Bioorthogonal affinity purification of direct kinase substrates. *Journal of the American Chemical Society, 127*(15), 5288–5289. https://doi.org/10.1021/ja050727t.