
Published paper

Tee, K.L and Wong, T.S (2013) Polishing the craft of genetic diversity creation in directed evolution. Biotechnology Advances, 31 (8). 1707 - 1721. Doi: 10.1016/j.biotechadv.2013.08.021

Biotechnology Advances 31 (2013) 1707–1721

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Polishing the craft of genetic diversity creation in directed evolution[☆]

Kang Lan Tee [a], Tuck Seng Wong [b,*]

[a] Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, England, United Kingdom
[b] ChELSI Institute, Department of Chemical and Biological Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, England, United Kingdom

ARTICLE INFO

Article history:
Received 28 May 2013
Received in revised form 31 August 2013
Accepted 31 August 2013
Available online 6 September 2013

Keywords:
Directed evolution
Genetic diversity
Random mutagenesis
Site-directed mutagenesis
DNA recombination
Synthetic biology
Genome engineering
Metabolic engineering

ABSTRACT

Genetic diversity creation is a core technology in directed evolution where a high quality mutant library is crucial to its success. Owing to its importance, the technology in genetic diversity creation has seen rapid development over the years and its application has diversified into other fields of scientific research. The advances in molecular cloning and mutagenesis since 2008 were reviewed. Specifically, new cloning techniques were classified based on their principles of complementary overhangs, homologous sequences, overlapping PCR and megaprimers and the advantages, drawbacks and performances of these methods were highlighted. New mutagenesis methods developed for random mutagenesis, focused mutagenesis and DNA recombination were surveyed. The technical requirements of these methods and the mutational spectra were compared and discussed with references to commonly used techniques. The trends of mutant library preparation were summarised. Challenges in genetic diversity creation were discussed with emphases on creating "smart" libraries, controlling the mutagenesis spectrum and specific challenges in each group of mutagenesis methods. An outline of the wider applications of genetic diversity creation includes genome engineering, viral evolution, metagenomics and a study of protein functions. The review ends with an outlook for genetic diversity creation and the prospective developments that can have future impact in this field.

Contents

Abbreviations: 8-oxo-dGTP, 8-Oxo-2′-deoxyguanosine-5′-triphosphate; AAV, Adeno-associated virus; ABI-REC, Asymmetric Bridge PCR with Intramolecular Homologous Recombination; bp, Base pairs; CAST, Combinatorial Active-Site Saturation Test; CPEC, Circular Polymerase Extension Cloning; CRP, cAMP receptor protein; $dA^{TESE}TP$, 7-Deaza-7-(triethylsilylethynyl)deoxyadenosine triphosphate; DGRs, Diversity-generating retroelements; dITP, 2′-Deoxyinosine 5′-triphosphate; DNA, Deoxyribonucleic acid; dNTP, Deoxyribonucleotide triphosphate; dPTP, 2′-Deoxy-P-nucleoside-5′-triphosphate; dsDNA, Double-stranded DNA; DuARCheM, Dual Approach to Random Chemical Mutagenesis; EMP, Exponential Megapriming PCR; EMS, Ethyl methane sulfonate; enoyl ACP reductase, Enoyl-acyl carrier protein reductase; epPCR, Error-prone polymerase chain reaction; epRCA, Error-prone rolling circle amplification; GOI, Gene of interest; GST, Glutathione-S-transferase; InDel, Insertion and deletion; ITCHY, Incremental Truncation for the Creation of Hybrid Enzymes; KF, Klenow fragment; MAP, Mutagenesis Assistant Program; MEGAWHOP, Megaprimer PCR of Whole Plasmid; MGS, Mutation Generation System; MLF-SDM, Megaprimed and Ligase-Free PCR-based Method for Site-Directed Mutagenesis; NEB, New England Biolabs; NiDE, Nicking DNA Endonuclease; NRR, Non-homologous random recombination; nt, Nucleotide; OLTA, OverLap extension PCR and TA cloning; OSCARR, One-pot Simple Methodology for Casette Randomization and Recombination; pBpa, *p*-Benzoylphenylalanine; PCR, Polymerase chain reaction; PERMUTE, PERMutation Using Transposase Engineering; PFLF-MSDM, Phosphorylation-Free and Ligase-Free PCR-based Method for Multiple SDM; PLICing, Phosphorothioate-based Ligase-Independent Gene Cloning; PPCP, PCR Production of Circular Plasmid; PS, Phosphorothioate; PTRec, Phosphorothioate-based DNA Recombination; RCA, Rolling circle amplification; REs, Restriction enzymes; RF cloning, Restriction-Free cloning; RGEN, RNA-guided Endonuclease; SDM, Site-directed mutagenesis; SEFC, Seamless Enzyme-Free Cloning; SeSaM, Sequence Saturation Mutagenesis; SHIPREC, Sequence Homology-Independent Protein RECombination; SLiCE, Seamless Ligation Cloning Extract; SPRINP, Single-Primer Reactions In Parallel; ssDNA, Single-stranded DNA; StEP, Staggered Extension Process; STRU-Cloning, Single-Tube Restriction-based Ultrafiltration Cloning; TaGTEAM, Targeting Glycosylases To Embedded Arrays for Mutagenesis; TALENs, Transcription activator-like effector nucleases; TAM, Transcription-associated mutation; TIM, Transposon Integration mediated Mutagenesis; TMGS-PCR, Truncated Metagenomic Gene-Specific PCR; TPCR, Transfer-PCR; TriNEx, TriNucleotide EXchange; TRINS, Tandem Repeat Insertion; T_s, Transitions; T_v, Transversions; UDG, Uracil-DNA glycosylase; USERec, USER Friendly DNA Recombination; ZFN, Zinc finger nuclease.

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[*] Corresponding author. Tel.: +44 114 2227591; fax: +44 114 2227501.
E-mail address: t.wong@sheffield.ac.uk (T.S. Wong).

0734-9750/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.biotechadv.2013.08.021

1. Introduction

Directed evolution has emerged as a key enabling technology for tailoring or altering the properties of biomolecules (*e.g.*, proteins and nucleic acids) and of microorganisms to satisfy a wide range of biotechnological applications [*e.g.*, industrial biocatalysis, biotransformation, bioremediation and synthetic biology (Cobb et al., 2012)]. Rooted in the Darwinian theory of evolution, a typical directed evolution experiment encompasses iterative rounds of gene mutagenesis and phenotype selection through high-throughput screening, until the desired trait is attained (Bloom and Arnold, 2009).

Creating a good mutant library is arguably the most critical component in all directed evolution exercises and it requires a combination of the right mutagenesis method and an efficient cloning system. Methods of genetic diversity creation have previously been reviewed by various research groups (Bornscheuer and Kazlauskas, 2011; Shivange et al., 2009; Wong et al., 2006b). Nonetheless, the rapidly transforming field of molecular biology has fuelled creativity in scientists and we continue to see innovations in the way mutant libraries are prepared. For instance, new discoveries or better understanding of the underlying mechanisms of enzymes (*e.g.*, recombinase) and genetic systems (*e.g.*, transposition) have expanded the systems and methodologies used in mutagenesis. Advancement in cloning technology has led to simplification of the 2-step gene mutagenesis and cloning into a 1-step protocol.

In this review, we would like to provide a critical update of the cloning techniques and the genetic diversity creation methods developed for mutant library preparation over the past six years (since 2008). Specifically, the review summarises new cloning strategies that attempt to improve the conventional restriction–ligation cloning method to make it more amendable to mutant library creation. This is followed by an update of the methodologies in random mutagenesis, focused mutagenesis and DNA recombination, as well as the challenges these methods address by comparisons to more widely applied methods [*e.g.*, error-prone polymerase chain reaction (epPCR), QuikChange mutagenesis and DNA shuffling]. This update would provide a useful guide to both new and experienced directed evolutionists when developing strategies in mutant library creation. Importantly, the method comparison allows us to identify current key challenges. Mutant libraries have now seen applications beyond protein engineering. This review will survey its wider applications and conclude with a perspective on the future developments in the field of genetic diversity creation.

2. Cloning mutant libraries

In virtually all directed evolution campaigns, experimental work commences with molecular cloning of the gene of interest (GOI) into a vector for subsequent gene mutagenesis or into an expression vector for protein synthesis in an appropriate host organism (*e.g.*, *Escherichia coli*, *Bacillus subtilis*, *Pichia pastoris* and *Saccharomyces cerevisiae*). Traditional PCR-based gene mutagenesis methods (*e.g.*, epPCR and DNA shuffling) also require cloning of the mutagenized genes.

Conventionally, directional gene cloning relies on creating sticky ends (or cohesive ends) on both ends of an insert using a pair of type II restriction enzymes (REs), followed by joining the digested insert with a recipient vector pre-treated with the same pair of REs using a DNA ligase (Fig. 1A). Despite being a technique still widely employed in many research laboratories for cloning a GOI, this lengthy and time-consuming process has its challenges for cloning large mutant libraries. Incomplete restrictive digestion and poor ligation efficiency, for instance, reduce cloning efficiency. Further, suitable unique restriction sites might not be readily available and the addition of restriction sites might introduce undesired extra amino acids in the resultant recombinant protein. To overcome some or all of the aforementioned drawbacks, new ideas have been proposed and some have further been developed into commercial kits. These recent cloning methods are based on four strategies depicted in Fig. 1: (1) complementary overhangs, (2) homologous sequences, (3) overlapping PCR and (4) megaprimers.

2.1. Molecular cloning based on complementary overhangs

Among the 4 strategies, cloning based on complementary overhangs (Fig. 1A) most resembles the conventional restriction–ligation cloning method. It varies from the conventional cloning in the ways complementary overhangs between the gene insert and vector were generated. Four out of the 5 recently reported methods that used this strategy bypass the use of Type II REs, which typically generates 2–4 bp overhangs on both ends of an insert. One of the methods replaced this step by two parallel asymmetric PCRs of the GOI; one had excess reverse primer with tailing bases and the other had excess forward primer with tailing bases. As such, single-stranded DNA (ssDNA) was produced in each asymmetric PCR. ssDNAs from both PCRs were pooled and annealed to form a double-stranded fragment bearing overhangs at both ends that corresponded to the restriction overhangs of cloning vector (Wang et al., 2009a). This method however remained dependent on REs which were used to prepare the vector. Shinomiya et al. demonstrated unidirectional cloning by cleaving two distinct cloning sites with a single engineered zinc finger nuclease (ZFN) for both the GOI and vector (Mori et al., 2009; Shinomiya et al., 2011). The ZFN recognizes and cleaves a 26-bp DNA target site, generating a 2-nucleotide (nt) overhang (Shinomiya et al., 2011). Despite the replacement of RE either in GOI preparation or in both GOI and vector preparation, both methods above used DNA ligase to covalently link GOI with recipient vector,

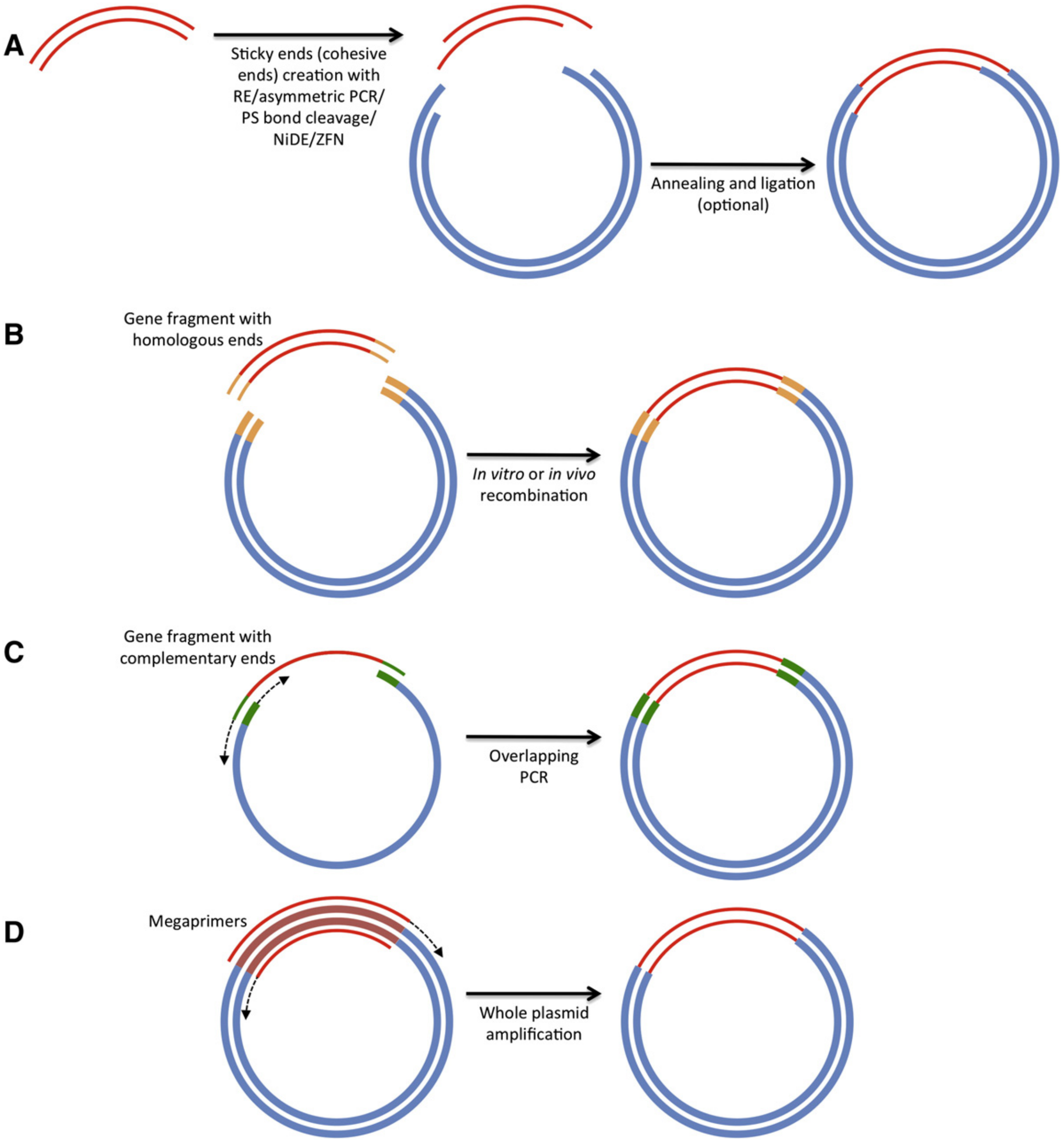

Fig. 1. Principles of molecular cloning: (A) complementary overhangs, (B) homologous sequences, (C) overlapping PCR and (D) megaprimers.

similar to that in conventional cloning. Phosphorothioate-based Ligase-Independent Gene Cloning (PLICing), reported by Blanusa et al., is a chemical-based method that bypasses both RE and DNA ligase (Blanusa et al., 2010). In PLICing, vector and GOI were amplified with synthetic oligonucleotides carrying multiple phosphorothioate (PS) linkages. The PS bond substitutes a sulfur atom for a non-bridging oxygen in the phosphate backbone of an oligonucleotide. After PCR, the PS bonds were cleaved with iodine/ethanol solution producing single-stranded overhangs. In another similar approach, termed NiDE, Yang et al. utilized a single Nicking DNA Endonuclease (*i.e.*, Nt.BbvCI) to create single-stranded overhangs on both insert and recipient vector (Yang et al., 2010). With the long complementary overhangs generated in PLICing (12-nt) and NiDE (14-nt), the GOI and vector form stable nicked plasmid DNA that can be directly transformed into *E. coli* without the use of DNA ligase. Further to the methodological improvements, technical modification was also adopted to achieve higher throughput in sticky end cloning. In Single-Tube Restriction-based Ultrafiltration Cloning (STRU-Cloning), Bellini et al. employed a centrifugal filter unit with membrane of suitable cut-off to remove small and unwanted DNA fragments created during restrictive digestion of plasmid or insert (Bellini et al., 2011). Despite being almost identical to conventional cloning technique, this approach avoided the time-consuming agarose gel electrophoresis that often results in low DNA recovery.

2.2. Molecular cloning based on homologous sequences

Methods that use homologous sequences (Fig. 1B) eliminate the need of complementary overhangs between the GOI and vector. This group of methods does not use REs or ligases but instead utilizes DNA recombination *in vitro* or *in vivo*. Seamless Ligation Cloning Extract (SLiCE) described a cloning approach using bacterial cell extract to assemble DNA fragments into recombinant DNA molecules in a single *in vitro* recombination reaction (Zhang et al., 2012). The authors showed that bacterial cell extract derived from standard laboratory *E. coli* K12 strains (*e.g.*, JM109 and DH10B) can be used to perform *in vitro* recombination and cloning efficiency was further enhanced using cell extract derived from a modified DH10B strain that expressed λ prophage Red recombination system. Highest cloning efficiency was achieved with an optimal homology length of 52 bp. The same principle has also proven to work *in vivo* (Abou-Nader and Benedik, 2010). Linearized vector was co-transformed into *E. coli* MB4091 strain [DH10B (pKD46)] together with the insert that shared homology flanking. pKD46 plasmid carries the λ *red* and *gam* genes that are expressed from the arabinose inducible *pBAD* promoter. Recombination rate was significantly enhanced by using dephosphorylated vector and gene fragment with 194- and 269-bp homologies on each end. Both methods using this strategy required long sequences of flanking homology, which can increase the cost of primers and possibly increase the difficulty of PCR during GOI amplification. Seamless Enzyme-Free Cloning (SEFC) is another *in vivo* method based on homologous sequences, in which PCR product and linear vector sharing homologous ends were co-transformed into *E. coli* XL10-Gold (Zhu et al., 2010). Contrary to the previous two techniques, the authors showed that a short homologous end of 15 bp was sufficient for creating recombinant DNA molecules, regardless of the homologous end being 5′-protruding, 3′-protruding or blunt (Zhong et al., 2013; Zhu et al., 2010).

2.3. Molecular cloning based on overlapping PCR

Circular Polymerase Extension Cloning (CPEC) is a one-step cloning procedure based on overlapping PCR (Quan and Tian, 2009). Following gene amplification, restrictive digestion and ligation were replaced by PCR to construct recombinant DNA molecules using linearized vector and insert that shared overlapping regions at both ends (Fig. 1C). For individual gene cloning, overlapping region of 25–27 nt and a single round of CPEC would suffice. CPEC was also successfully used for constructing a combinatorial library and for multi-component assembly (Quan and Tian, 2009), making it an attractive alternative to conventional cloning.

2.4. Molecular cloning based on megaprimer

Cloning based on megaprimer strategy (Fig. 1D) is, in principle, similar to QuikChange mutagenesis. Transfer-PCR (TPCR) presented by Erijman et al. amplified GOI with two primers containing 5′-sequences corresponding to the integration sites in the recipient vector (Erijman et al., 2011). Following the initial amplification stage, the generated intermediate PCR products served as megaprimers for linear amplification of whole plasmid. The concept of TPCR is identical to Megaprimer PCR of Whole Plasmid (MEGAWHOP cloning) (Miyazaki, 2011) and Restriction-Free cloning (RF cloning) (Unger et al., 2010). Exponential Megapriming PCR (EMP) cloning performs two PCR stages in a one-pot reaction (Ulrich et al., 2012). In the first stage, the gene was amplified to introduce only one vector-integration site at one end. In the second stage, the gene was integrated into the recipient vector *via* PCR. After plasmid circularization by T4 DNA ligase, the recombinant DNA molecules were ready for transformation. EMP was shown to have higher efficiency compared to RF cloning, especially for long inserts above 2.5 kb (Ulrich et al., 2012). Both strategies of overlapping PCR and megaprimer rely on the ability of polymerases to amplify long sequences accurately and efficiently, and are facilitated by the advancement and discovery of new polymerases in recent years.

2.5. Trend in molecular cloning

A comparison of the aforementioned molecular cloning techniques (Table 1) revealed that most methods are RE-free and ligase-independent. The transformation efficiencies reported provide a useful overview when selecting a cloning method but are not absolute comparisons since cloning efficiency is dependent on various factors (Yoshida and Sato, 2009), including transformation method used (electroporation or chemical method), bacterial cell competency, vector size, insert length, length of complementary sequences, and homology length which cannot be accounted for here. It is interesting to note that combination of the methods, overlapping PCR and homologous recombination in standard *E. coli* DH5α strain, was demonstrated in Asymmetric Bridge PCR with Intramolecular Homologous Recombination (ABI-REC) (Bi et al., 2012). ABI-REC requires only one flanking homologous sequence and high efficiency was reported using a relatively short homology length of 25 bp.

2.6. Improvements to cloning vectors

Further to the new cloning strategies, parallel efforts were made to improve cloning vectors. *ibsC* gene, which encodes a 19-residue toxin from *E. coli* K-12, was introduced into cloning vectors to facilitate

Table 1
Molecular cloning techniques reported since 2008.

Method	Strategy	RE and ligase requirement	*E. coli* strain	Transformation method	Transformation efficiency	Reference
Asymmetric PCR	Complementary overhangs	RE and ligase required	JM109	NR	9×10^3 cfu/μg DNA	Wang et al. (2009a)
ZFN	Complementary overhangs	RE-free	DH5α	NR	NR	Mori et al. (2009), Shinomiya et al. (2011)
PLICing	Complementary overhangs	RE- and ligase-free	XL10-Gold	Chemical method	8×10^5 cfu/μg DNA	Blanusa et al. (2010)
NiDE	Complementary overhangs	RE- and ligase-free	DH5α	Chemical method	1.2×10^4 cfu/μg vector	Yang et al. (2010)
STRU-Cloning	Complementary overhangs	RE and ligase required	DH5α	NR	NR	Bellini et al. (2011)
SLiCE	Homologous sequences	RE- and ligase-free	DH10B	Electroporation	7.7×10^8 cfu/μg vector	Zhang et al. (2012)
In vivo recombination	Homologous sequences	RE- and ligase-free	MB4091 [DH10B (pKD46)]	Electroporation	8.5×10^4 cfu/pmol insert	Abou-Nader and Benedik (2010)
SEFC	Homologous sequences	RE- and ligase-free	XL10-Gold	Chemical method	NR	Zhu et al. (2010)
CPEC	Overlapping PCR	RE- and ligase-free	GC5α	Chemical method	5.6×10^7 cfu/μg insert	Quan and Tian (2009)
TPCR	Megaprimers	RE- and ligase-free	DH5α	Chemical method	NR	Erijman et al. (2011)
RF cloning	Megaprimers	RE- and ligase-free	DH5α	NR	NR	Unger et al. (2010)
EMP cloning	Megaprimers	RE-free	DH5α	Chemical method	NR	Ulrich et al. (2012)
ABI-REC	Overlapping PCR + homologous sequences	RE- and ligase-free	DH5α	Chemical method	NR	Bi et al. (2012)

selection of clones harbouring vector with a gene insert (Mok and Li, 2013). Another toxic *ccdB* gene was also a popular choice for selection of positive clones (Wang et al., 2013a; Scholz et al., 2013). A high cloning efficiency of 95–100% (*i.e.*, <5% background) was reported using these "detox cloning" systems (Mok and Li, 2013; Scholz et al., 2013; Wang et al., 2013a). Antibiotic resistance genes (*e.g.*, β-lactamase, neomycin phosphotransferase II) are commonly used as selection marker for plasmid-bearing bacteria. A *mfabI* gene that encodes a mutant form (G93V) of FabI (enoyl ACP reductase) was demonstrated recently as an efficient selection marker when *E. coli* cells were grown in the presence of triclosan (Jang and Magnuson, 2013). The lack of unique restriction sites within a gene insert was earlier identified as a possible challenge to directional cloning. Instead of using a RE-free cloning method to overcome this problem, Kielkowski et al. developed an alternative strategy to protect internal restriction sites against RE cleavage by incorporating 7-deaza-7-(triethylsilylethynyl) deoxyadenosine triphosphate ($\mathrm{dA^{TESE}TP}$) into DNA during gene amplification *via* PCR (Kielkowski et al., 2013). The silylethynyl-protected DNA was resistant to RE cleavage, whereas the restriction sites within PCR primers were still cut by their corresponding REs (Kielkowski et al., 2011). These latest developments would mean that directed evolutionists are provided with more choices (*e.g.*, vectors, selection methods, cloning strategies) while preparing a gene library.

3. Genetic diversity creation

An ideal gene library for directed evolution would satisfy five requirements (Table 2). Firstly, the library should be complex enough to contain rare beneficial mutations. Secondly, the library should encode for mostly functional and properly folded proteins. Thirdly, the library contains mostly unique gene sequences with none or minimal genotype duplication. Fourthly, the mutational spectrum of the library can be adjusted to populate certain types of amino acid substitutions, depending on the properties to be evolved. Fifthly, the library can be created easily and cost effectively. Satisfying these requirements would increase the chances of identifying rare beneficial mutations and reduce screening effort (Table 2). Broadly, methods for creating genetic diversity can be classified into three categories: random mutagenesis, focused mutagenesis and DNA recombination. In each category, many methods have been reported with varying degrees of technical difficulty, cost effectiveness and resultant library quality.

3.1. Random mutagenesis

Random mutagenesis is widely used in directed evolution for quick and easy gene library preparation. Popular methods of choice include *Taq* polymerase-based epPCR (*e.g.*, $MnCl_2$ and imbalanced dNTP concentration), Mutazyme, nucleotide analogous (*e.g.*, 8-oxo-dGTP, dPTP and dITP), mutator strain (*e.g.*, XL1-Red) and chemical mutagens [*e.g.*, ethyl methane sulfonate (EMS) and hydroxylamine] (Wong et al., 2006b). Commercial kits are available for most of them. In recent years, we have witnessed further simplification of these methods. MegAnneal (Pai et al., 2012) and PCR Production of Circular Plasmid [PPCP; (Le et al., 2013)] combined epPCR and cloning *via* the megaprimer strategy discussed earlier into a single step (Erijman et al., 2011; Miyazaki, 2011; Unger et al., 2010). Another variation used was the cloning of epPCR libraries *via* overlapping PCR, in a similar fashion as in CPEC (Quan and Tian, 2009). These strategies simplify and expedite the 2-step mutagenesis and cloning of library creation but they do not change the diversity achieved or mutational spectrum control in currently existing mutagenesis methods.

A good random mutagenesis method seeks to cover all nucleotide substitutions equally and to achieve 3 consecutive nucleotide substitutions to target all amino acid changes (Wong et al., 2007). An ideal random mutagenesis method would allow equal occurrence of all four transitions ($\mathrm{T_s}$: AT → GC and GC → AT) and eight transversions ($\mathrm{T_v}$: AT → TA, AT → CG, GC → CG and GC → TA), with a probability of 16.67% for each nucleotide substitution pair and a T_s/T_v ratio of 0.5 (Table 3) (Wong et al., 2006a,b). Further, there should be no insertion and deletion (InDel). In 2009, Savilahti and coworkers conducted a comparative evaluation of mutational spectra from gene libraries prepared using various random mutagenesis methods by sequencing 11,500–21,500 nucleotides (Rasila et al., 2009). These statistics serve as a good benchmark and are included in the preferential nucleotide substitution comparison of the new random mutagenesis methods reported here (Table 3).

3.1.1. Variations to previous random mutagenesis methods

Many recently published methods are variations of existing protocols. Dual Approach to Random Chemical Mutagenesis (DuARCheM), for example, is a two-stage method (Mohan and Banerjee, 2008) where GOI was first randomly mutated *via in vivo* chemical mutagenesis with EMS. Subsequently, treated genes were isolated and cloned into untreated expression vectors to avoid mutations in the plasmid backbone. Despite having a T_s/T_v of 0.7 which is close to the ideal T_s/T_v value of 0.5, the GC → CG transversion is heavily overrepresented (Table 3). Minamoto et al. attempted PCR using heavy water (D_2O) as solvent, in the absence or presence of $MnCl_2$ (Minamoto et al., 2012). *Taq* DNA polymerase exhibited 8-fold higher replication error in D_2O ($\sim 1.2 \times 10^{-3}$ errors/bp) compared to that in H_2O (1.5×10^{-4} errors/bp). This error rate was further increased to $\sim 1.8 \times 10^{-3}$ errors/bp when PCR was done in D_2O in the presence of $MnCl_2$. This was the first instance where D_2O was used as a mutagen in random mutagenesis. Although the mechanism of D_2O-induced mutagenesis is yet to be

Table 2
Five requirements of an ideal gene library for directed evolution.

Requirements	Implication for genetic diversity creation methods	Implications for screening
1. Complex library to contain rare beneficial mutations	• No/minimal wildtype sequence • No/minimal mutational bias • Consecutive nucleotide substitutions or codon-based mutagenesis	• Smaller library to screen
2. Encode for mostly functional and properly folded proteins	• Moderate mutation rate • No/minimal prematurely truncated genes due to introduction of stop codons or frameshift mutations (*e.g.*, insertion, deletion) • No/minimal structurally-disrupting mutations (*e.g.*, introducing Gly/Pro in helix)	• Smaller library to screen
3. Contains mostly unique gene sequences with none or minimal genotype duplication	• No/minimal mutational hotspots or preferential sites • Mutations randomly distributed across the entire gene	• Avoid screening identical clones (*i.e.*, more effective screening)
4. Possibility of populating certain amino acid substitutions depending on property to be evolved	• Adjustable mutational spectrum	• Smaller library to screen
5. Easy and cost-effective preparation	• Minimal DNA manipulation • No expensive kits/enzymes/chemicals • Minimal number of oligonucleotides required • Library can be created within a shorter time frame	• Screening can commence sooner

Table 3
Random mutagenesis methods reported since 2008. Preferential nucleotide substitutions (*i.e.*, probability > 16.7%) were in bold.

Methods	Transitions		Transversions				InDel	T_s/T_v	Reference
	AT →GC	GC →AT	AT →TA	AT →CG	GC →CG	GC →TA			
Ideal method	16.7	16.7	16.7	16.7	16.7	16.7	0.0	0.5	Wong et al. (2006a), Wong et al. (2006b)
Taq/$MnCl_2$ and imbalanced dNTP concentration	38.1	19.0	27.8	4.0	0.0	6.3	4.8	1.5	Rasila et al. (2009)
Taq/8–oxo–dGTP and dPTP	65.4	27.5	1.6	4.4	0.0	0.0	1.1	15.5	Rasila et al. (2009)
Mutazyme/Amplicon	14.0	32.5	15.1	2.3	5.8	22.1	8.3	1.0	Rasila et al. (2009)
Mutazyme/Cycle	17.1	25.7	28.6	2.9	0.0	14.3	11.4	0.9	Rasila et al. (2009)
XL1–Red	0.0	60.0	10.0	0.0	0.0	0.0	30.0	6.0	Rasila et al. (2009)
NH_2OH–HCl	15.4	76.9	7.7	0.0	0.0	0.0	0.0	12.0	Rasila et al. (2009)
DuARChEM	13.5	26.9	1.9	5.8	42.3	9.6	0.0	0.7	Mohan and Banerjee (2008)
Taq/D_2O	50.0	37.5	12.5	0.0	0.0	0.0	0.0	7.0	Minamoto et al. (2012)
Taq/D_2O and $MnCl_2$	60.0	12.0	8.0	8.0	4.0	8.0	0.0	2.6	Minamoto et al. (2012)
dITP and endonuclease V	38.4	39.7	4.6	8.6	4.6	4.0	0.0	3.6	Wang et al. (2013c)
epRCA with Cre/*loxP* recombination	6.7	86.7	0.0	0.0	6.7	0.0	0.0	13.9	Huovinen et al. (2011)
SeSaM–Tv –II	15.2	18.3	1.8	1.8	33.9	29.1	0.0	0.5	Mundhada et al. (2011)
TaGTEAM	6.1	10.2	18.4	0.0	14.3	26.5	24.5	0.3	Finney–Manchester and Maheshri (2013)

elucidated, it was demonstrated that PCR yield was not affected in D_2O. Mutational spectrum of the method was heavily skewed towards transition, with a T_s/T_v ratio higher than that of epPCR with $MnCl_2$ and imbalanced dNTP concentration (Table 3). Wang et al. created a gene library by combining PCR with 2′-deoxyinosine 5′-triphosphate (dITP) and DNA fragmentation with endonuclease V (Wang et al., 2013c), which is a deoxyinosine 3′-endonuclease (Cao, 2012). dITP is a purine analogue that preferentially base-pairs with T and C (Wong et al., 2008). The use of endonuclease V allows nucleotide substitutions to be shuffled and recombined, thereby increasing mutation frequency. As with numerous other methods, Wang's method preferred transitions over transversions owing to the base-pairing property of dITP.

Lamminmaki group improved the efficiency of previously reported error-prone rolling circle amplification [epRCA; (Fujii et al., 2004)] by DNA concatemer resolution through Cre/*loxP* recombination, as opposed to cutting the DNA concatemer to plasmid-sized fragments with unique RE and circularizing the fragments through self-ligation (Huovinen et al., 2011). However the method is the most transition biassed among the new methods in Table 3. The same group also demonstrated selective amplification of nascently synthesized circular DNA carrying desired mutation by ϕ29 DNA polymerase, through uracil-DNA glycosylase (UDG) treatment of uracil-containing parental strand (Huovinen et al., 2012).

Liu et al. proposed mutagenizing GC-rich genes using sodium bisulfite as chemical mutagen (Liu et al., 2009). Sodium bisulfite catalyzes specific deamination of unmethylated cytosine to uracil. The mutational spectrum of the method was not shown but likely favours transitions due to replacement of cytosine by uracil.

All methods discussed so far can only statistically target single nucleotide within a codon, thus unable to achieve substitutions to all 19 other amino acids. A recently published method which can target consecutive nucleotide substitution is SeSaM-Tv-II (Mundhada et al., 2011), an improved version of Sequence Saturation Mutagenesis (SeSaM) (Wong et al., 2004, 2005, 2008). Up to 37% of the sequenced clones carried consecutive nucleotide substitutions (*e.g.*, T_vT_v), a mutational pattern unobtainable with epPCR or its derivatives. Among them, clones with three consecutive nucleotide substitutions were also reported (*e.g.*, $T_vT_sT_s$). SeSaM has been successfully applied for the evolution of *Bacillus gibsonii* alkaline protease (Martinez et al., 2013) and *Yersinia mollaretii* phytase (Shivange et al., 2012).

A careful examination of Table 3 revealed that most methods are still bias towards transitions ($T_s/T_v > 0.5$). XL1-Red and Mutazyme resulted in high percentage of InDel. Further, transversion of AT → CG is under-represented, which could possibly be solved by SeSaM-Tv-II through DNA fragmentation with dATPαS/dTTPαS followed by elongation with dPTP/dITP. However, this remains to be experimentally proven.

3.1.2. Novel concept in random mutagenesis

Maheshri group recently reported Targeting Glycosylases To Embedded Arrays for Mutagenesis (TaGTEAM), which is distinctly different from epPCR (Finney-Manchester and Maheshri, 2013). The method was designed for targeted *in vivo* mutagenesis in yeast (Finney-Manchester and Maheshri, 2013). They fused yeast 3-methyladenine DNA glycosylase (Mag1) to a tetR DNA-binding domain and localized it to an array of binding sites in *S. cerevisiae* (*i.e.*, array of tet operator sequences). Mag1 creates abasic sites in the region surrounding its binding sites. These Mag1-induced damages generate intermediates for homologous recombination repair in an error-prone manner. Although this method showed higher occurrence of transversions ($T_s/T_v < 0.3$), the deletion percentage was very high (24.5%).

3.1.3. Transposon-based random mutagenesis

An exciting development in the field is the emergence of transposition-based methods (Table 4). These methods use MuA transposase for random domain/tag/multiple amino acid insertion (Edwards et al., 2008; Hoeller et al., 2008), random protein truncation, random nucleotide triplet substitution (Baldwin et al., 2008; Daggett et al., 2009; Liu and Cropp, 2012) or random circular permutation (Mehta et al., 2012), depending on the design of mini-Mu transposon (Table 4). Commercial kits of similar principles are also available from Thermo Scientific (*e.g.*, Mutation Generation System and Stop Generation System). Despite being capable of generating more complex gene libraries compared to methods listed in Table 3, they often involve extensive DNA manipulation (*e.g.*, multiple transformation, restrictive digestion and ligation steps), thus reducing overall efficiency. Further, transposition can occur in the vector backbone (*i.e.*, GOI is not mutated) or result in frameshift (*i.e.*, non-functional sequences). The transposon integration efficiency is also affected by various factors, such as orientation, site preference and transposon size. That being said, the development of these methods is a breakthrough in the field. Methods presented by Cropp and coworkers, for instance, permit mutagenizing single or multiple codons throughout a protein sequence, by cleverly incorporating a β-lactamase assay to select for in-frame transposition (Daggett et al., 2009; Liu and Cropp, 2012). This method can potentially achieve any amino acid substitution. For example, a random codon change to an amber stop codon (TGA) would allow incorporation of unnatural amino acids (Liu and Schultz, 2010; Wang et al., 2012) at random position (Daggett et al., 2009). To illustrate the applicability of their method, Cropp and coworkers inserted *p*-benzoylphenylalanine (pBpa), an unnatural photo-crosslinking amino acid, randomly in glutathione-S-transferase (GST) (Daggett et al., 2009). In light of the technical requirements, transposon-based methods suit more advanced users and experienced directed evolutionists.

3.1.4. Altered target sequence length in random mutagenesis

Methods discussed so far are largely limited to side-chain substitution (point mutation). Tawfik and coworkers introduced Tandem Repeat Insertion (TRINS) that allows generation of tandem repeats of random fragments of GOI *via* rolling circle amplification (RCA) and concurrent incorporation of these repeats into GOI (Kipnis et al., 2012). Using OverLap extension PCR and TA cloning (OLTA), Fujii et al. modified ZFN by creating repeating copies of DNA-binding zinc fingers (Fujii et al., 2013). Should the gene length be kept (almost) the same in a gene library (*e.g.*, libraries created using epPCR, Mutazyme, SeSaM, TriNEx, or Codon Scanning Mutagenesis)? Or is it wiser to have variable gene length in a gene library (*e.g.*, libraries created using TRINS or OLTA)? This remains debatable, but likely depends on the protein of interest and the properties to be evolved. As illustrated by Fujii et al., tandem repeats of DNA-binding motifs help to improve DNA-binding affinity (Fujii et al., 2013).

3.2. Focused mutagenesis

Improved protein variants found after screening allow us to identify beneficial mutations. In some cases, the mutated positions or their neighbouring positions are subject to saturation mutagenesis, with the aim of further enhancing the protein property or understanding structure–function relationships. In rational designs of protein, key residues are often identified for randomization after examining the molecular structure of the target protein. Therefore, simple and robust techniques for site-directed/multisite-directed/cassette mutagenesis are highly desirable. Table 5 summarises focused mutagenesis methods reported since 2008 and compares these methods based on (1) the quality of the resultant mutant library and (2) the experimental requirements/technical difficulty.

3.2.1. QuikChange derivatives

Since 2008, multiple groups improved the QuikChange method of linear whole plasmid amplification (Fig. 2). QuikChange, the most frequently used method for focused mutagenesis, bypasses the cumbersome RE digestion and ligation (Fig. 2A). The method, however, is inherited with two drawbacks owing to its primer design. First, the mutagenic primers are complementary to each other, therefore favouring primer-dimer formation (in opposed to primer–template hybridization) and lowering efficiency. Second, PCR fails when mutagenic primers anneal to newly synthesized “nicked” daughter DNAs.

Single-Primer Reactions In Parallel (SPRINP) (Fig. 2B) uses two parallel PCRs with only one mutagenic primer in each PCR to circumvent the primer-dimer problem (Edelheit et al., 2009). After the first amplification stage, PCR products were combined and additional PCR cycles were run. SPRINP is almost identical to another 2-stage protocol published (Wang and Malcolm, 1999), with the exception of DNA polymerase used. By using a pair of partially overlapping mutagenic primers (Fig. 2C), Liu and Naismith achieved higher QuikChange efficiency (Liu and Naismith, 2008). Similar strategy has also been reported by Reymond and coworkers previously (Zheng et al., 2004).

In a 2-stage protocol proposed independently by Reetz group (Sanchis et al., 2008) and Tseng et al. [Megaprimed and Ligase-Free PCR-based Method for Site-Directed Mutagenesis, MLF-SDM; (Tseng et al., 2008)], a sequence within the gene-harbouring plasmid was amplified using a mutagenic primer and a silent primer that were non-overlapping (Fig. 2D). The PCR products from the first stage served as megaprimers for whole plasmid amplification in the second stage. Tseng et al. extended their megaprimer-based 2-stage procedure for simultaneous mutation of up to six distal sites (Phosphorylation-Free and Ligase-Free PCR-based Method for Multiple SDM, PFLF-MSDM; (Tseng

Table 4
Transposon-based random mutagenesis methods reported since 2008.

Method	Genetic diversity	Reading-frame selection	Resultant protein length (theoretical)	Transposase	Transposon	Reference
Mutation Generation System (MGS)	Random insertion of 15 bp	No	(X + 5) a.a.	MuA	Symmetrical mini-Mu carrying kanamycin resistance gene, flanked by MuA and NotI recognition sites	Thermo Scientific technical manual
Stop Generation System	Random protein truncation	No	Variable	MuA	Symmetrical mini-Mu carrying kanamycin or chloramphenicol resistance gene, flanked by MuA and NotI recognition sites	Thermo Scientific technical manual
Domain insertion	Random domain insertion of N a.a.	No	(X + N − 1) a.a.	MuA	Symmetrical mini-Mu carrying chloramphenicol resistance gene, flanked by MuA and MlyI recognition sites	Edwards et al. (2008)
TriNEx	Random nucleotide triplet substitution	No	X a.a.	MuA	Symmetrical mini-Mu carrying chloramphenicol resistance gene, flanked by MuA and MlyI recognition sites	Baldwin et al. (2008)
TIM	Random tag insertion of N a.a.	No	(X + N) a.a.	MuA	Symmetrical mini-Mu carrying kanamycin resistance gene, flanked by MuA and AarI recognition sites	Hoeller et al. (2008)
Codon scanning mutagenesis	Random nucleotide triplet substitution	Yes	X a.a.	MuA	Symmetrical mini-Mu carrying chloramphenicol resistance gene, flanked by MuA and MlyI recognition sites	Daggett et al. (2009)
Multi-codon scanning mutagenesis	Random single, double or triple nucleotide triplet substitution	Yes	X a.a.	MuA	Asymmetrical mini-Mu carrying VMA-intein and β-lactamase fusion gene, flanked by MuA recognition sites	Liu and Cropp (2012)
PERMUTE	Circular permutation	No	(X + 23) a.a.	MuA	Symmetrical mini-Mu carrying a stop codon, a terminator, an origin of replication, a kanamycin resistance gene, a promoter and a ribosome-binding site, flanked by MuA recognition sites	Mehta et al. (2012)

Table 5
Comparison of focused mutagenesis methods published since 2008.

Method	Library quality				Experimental requirements							Reference
	Single/multi site	Maximum number of sites	Efficiency (number of sites)	Possibility of insertion (efficiency)/deletion (efficiency)[a]	Primer length[b]	Number of primers for substitution[c]	Primer modification	Polymerase used[d]	Cloning with restriction enzymes	Type of DNA transformed	Parental template removal	
QuikChange lightning site-directed mutagenesis	Yes/no	1	85% (1)	Yes/yes	25–45 nt	2	None	PfuUltra high-fidelity	No	Nicked dsDNA	DpnI digest	Agilent manual (catalogue number 210518)
QuikChange Lightning Multi Site-Directed Mutagenesis	Yes/yes	5	55% (3)	NR/NR	25–45 nt	N	None (kinase required)	Pfu Fusion	No	Circular ssDNA	DpnI digest	Agilent manual (catalogue number 210514)
QuikChange derivative: SPRINP	Yes/no	1	100% (1)	Yes/NR	33–57 nt	2	None	Pwo Master	No	Nicked dsDNA	DpnI digest	Edelheit et al. (2009)
QuikChange derivative: Partially overlapping primers	Yes/yes	2	75% (2) 100% (1)	Yes (>75%)/ yes (100%)	34–56 nt	2 N	None	Pfu	No	Nicked dsDNA	DpnI digest	Liu and Naismith (2008)
QuikChange derivative: MLF-SDM	Yes/yes	2	90% (1)	NR/NR	22–40 nt	2	None	PfuTurbo	No	Nicked dsDNA	DpnI digest	Tseng et al. (2008)
QuikChange derivative: PFLF-MSDM	Yes/yes	6	40% (6) 80% (4)	NR/NR	22–23 nt	N	None	PfuTurbo	No	Nicked dsDNA	DpnI digest	Tseng et al. (2010)
QuikChange derivative: Ω-PCR	Yes/no	1	94–95% (1)	Yes (78–100%)/ yes (100%)	41–43 nt	2	None	PrimeSTAR Taq	No	Nicked dsDNA	DpnI digest	Chen et al. (2013)
QuikChange derivative: Q5 Site-Directed mutagenesis	Yes/no	1	>90% (1)	Yes (>90%)/ yes (>90%)	>23 nt	2	None (kinase required)	Q5	No	Circular dsDNA	DpnI digest	NEB manual (catalogue number E0554S)
QuikChange derivative: PFunkel	Yes/yes	4	70% (4) 100% (1)	NR/NR	25–26 nt	N + 1	5′-phosphorylation	PfuTurbo Cx Hotstart	No	Circular dsDNA	UDG and ExoIII degradation of uracil-containing ssDNA	Firnberg and Ostermeier (2012)
OmniChange	Yes/yes	5	100% (5)	NR/NR	25–36 nt	2 N	Phosphorothioate bond	Phusion	No	Nicked dsDNA	DpnI digest	Dennig et al. (2011)
OSCARR	Yes/no	1	95% (1)	NR/NR	18–50 nt	3	None	PfuPlus!	Yes	Circular dsDNA	Gel purification	Hidalgo et al. (2008)
In vivo method: *E. coli*	Yes/no	1	~40% (1)	NR/NR	~73 nt	1	None	Not required	No	Oligonucleotide	No template removal (Mutate a gene within genome)	Valledor et al. (2012)
In vivo method: yeast	Yes/yes	3	0.01% (3) 0.2% (2) 5% (1)	NR/NR	63 nt	N	None	Not required	No	Oligonucleotide and linear vector	No template removal	Pirakitikulr et al. (2010)
Partially *in vivo* method: *E. coli*	Yes/no	1	100% (1)	Yes (93%)/ yes (100%)	30–47 nt	4	None	Pfu	Yes	Linear dsDNA	No template removal (PCR from genomic DNA)	Wu et al. (2013)

[a] NR = Not reported.
[b] nt = Nucleotides.
[c] N = Number of sites.
[d] Polymerase used determines the time required for PCR and the accuracy of vector backbone amplification.

et al., 2010)). Compared to QuikChange Multi Site-Directed Mutagenesis, PFLF-MSDM does not require 5′-phosphorylated primers and ligation of mutated fragments. Further, mutagenic efficiency reported was higher than that of QuikChange method [66.5% for triple mutations, 45.4% for quadruple mutations and 32.4% for pentuple mutations; (Hogrefe et al., 2002)]. Ω-PCR, published very recently, is another single site mutagenesis method based on the use of megaprimers for whole plasmid amplification (Chen et al., 2013).

In Q5 Site-Directed Mutagenesis Kit commercialised by New England Biolabs (NEB), two non-overlapping primers (one mutagenic and one silent) are used for whole plasmid amplification (Fig. 2E). Subsequently, PCR product is subject to kinase-ligase-DpnI enzyme mix treatment to 5′-phosphorylate and circularise daughter DNAs and to remove parental templates.

In PFunkel (Firnberg and Ostermeier, 2012), a 5′-phosphorylated mutagenic primer was used to amplify uracil-containing ssDNA template and the mutant strand was circularized with *Taq* ligase (Fig. 2F). Complementary strand was subsequently synthesized by addition of a reverse primer. After removal of template DNA using UDG and endonuclease III, double-stranded mutant DNAs were transformed into *E. coli*. PFunkel was used to saturate four distal sites with an efficiency of 70%. The authors also demonstrated the saturation of every single codon of TEM-1 β-lactamase gene (287 codons) by using 287 mutagenic primers in a one-pot reaction with a low primer to template ratio. Compared to PFunkel, QuikChange is technically simpler since it does not require preparation of uracil-containing ssDNA template and 5′-phosphorylated mutagenic primer. Among the five QuikChange variations (Fig. 2B–F), methods of using partially overlapping mutagenic primers, NEB kit and PFunkel overcome both the drawbacks of original QuikChange.

3.2.2. Novel concepts in focused mutagenesis

Two novel concepts, OmniChange and One-pot Simple Methodology for Casette Randomization and Recombination (OSCARR), were introduced. OmniChange allows simultaneous saturation of five codons in a GOI (Dennig et al., 2011). Vector carrying GOI was amplified in parallel PCRs to generate fragments that cover the entire plasmid sequence. Each PCR was done using a pair of primers (one mutagenic and one silent) that contained PS linkages; the number of parallel PCRs required equals the number of sites to be simultaneously mutated/saturated. After parallel PCRs, PS linkages were cleaved with iodine/ethanol producing single-stranded overhangs. All DNA fragments were quantified, pooled, annealed, and transformed into *E. coli*. OmniChange achieved 100% mutagenic efficiency in an attempt to saturate 5 codons simultaneously using primers with NNK degenerate codons (N: A/T/G/C, K: G/T; 32 codons for the entire set of standard amino acids). The high mutagenic efficiency is likely due to the inherent nature of the method where DNAs can only be transformed into *E. coli* when all PCR fragments (therefore all mutated sites) anneal to form complete plasmids. An analysis of 48 clones revealed that an NNK coverage of 65.6–84.4% was feasible with OmniChange.

OSCARR is a 2-stage protocol developed for randomizing a preselected protein region (5–16 amino acids) (Hidalgo et al., 2008). In the first stage, a partial sequence of GOI was amplified using a spiked oligonucleotide (covering the region to be randomized) and a silent primer in an asymmetric PCR. In the second stage, a second silent primer was added to generate the full-length GOI. The authors further optimized their method for randomizing two protein regions. Owing to its "one-pot" nature and the use of spiked oligonucleotide, wildtype sequences or sequences with synonymous mutations were expected in the gene library; thereby reducing genetic diversity. Nonetheless, OSCARR allows randomizing a longer protein region in comparison to conventional QuikChange method that is mostly used for 3 nt substitutions (*i.e.*, a single codon). In contrast to all aforementioned methods, OSCARR library requires cloning using REs and DNA ligase.

3.2.3. In vivo methods for focused mutagenesis

In vivo methods represent a varied development from the prevalent *in vitro* mutagenesis methods. Valledor et al. electroporated single-stranded mutagenic oligonucleotide into *E. coli* cells that transiently expressed λ prophage recombination system encoded by *exo*, *bet* and *gam* genes to alter gene encoded in the genome (Valledor et al., 2012).

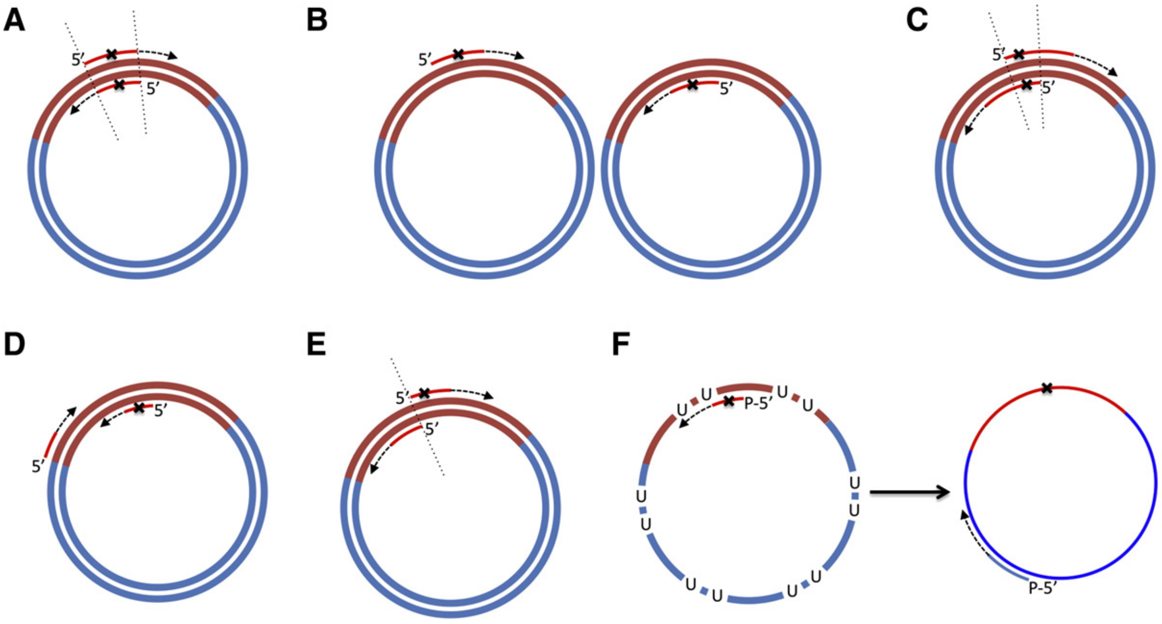

Fig. 2. Primer design in original QuikChange method and QuikChange-derivatives: (A) Complementary mutagenic primer pair used in original QuikChange method, (B) Two parallel PCRs using one mutagenic primer each, (C) Partially overlapping mutagenic primer pair, (D) PCR with one mutagenic primer and one silent primer. Intermediate PCR products serve as megaprimers for subsequent whole plasmid amplification. (E) Non-overlapping primers (one mutagenic and one silent) used in NEB kit, (F) 5′-phosphorylated mutagenic primer used in PFunkel. Complementary strand is subsequently synthesized by addition of a reverse primer.

Though convenient, the library contained high proportion of wildtype sequence due to low recombination efficiency. Unlike most other methods that utilize RE DpnI to degrade methylated parental plasmids, *in vivo* method does not separate mutant DNAs from the starting templates. In the same vein, Pirakitikulr et al. reported an *in vivo* mutagenesis in yeast, in which linear vector containing GOI was co-transformed with single-stranded mutagenic oligonucleotide (Pirakitikulr et al., 2010). Multiple sites could be mutated simultaneously, despite low efficiency. Another partially *in vivo* technique that involves more DNA manipulation was designed by Wu et al. (2013). This method involved (1) creating two PCR fragments in two separate PCRs, each with one mutagenic primer and one silent primer, (2) digesting the fragments individually with one unique RE, (3) ligating the fragments to a linearized vector to produce linear recombinant vector with two blunt ends which are homologous, and (4) transforming the linear vector into a standard *E. coli* DH5α for plasmid circularization *via* recombination. Although time-consuming, the method achieved high efficiency for nucleotide substitution, deletion and insertion.

3.3. DNA recombination

Gene library creation *via* DNA recombination (or DNA shuffling) is also widely adopted in molecular evolution. Contrary to random mutagenesis that accumulates mutations on a single parental gene, DNA recombination requires several gene sequences encoding proteins of similar function or a gene pool from random mutagenesis (*e.g.*, epPCR).

DNA shuffling, proposed originally by Stemmer, was a method for reassembly of genes from their random DNA fragments resulting in *in vitro* recombination (Stemmer, 1994). Briefly, a gene pool was subject to random fragmentation by DNaseI. Fragments of 10–50 bp were purified from agarose gel and reassembled in a primerless PCR. The product of primerless PCR was subsequently amplified in a second PCR using a pair of flanking primers. In an attempt to create a chimeric library between human and murine interleukin 1β (high DNA sequence identity between both genes; ClustalW2 score of 76), Stemmer obtained on average 1.9 crossovers per gene. Further, random mutations were introduced in a DNA shuffling process, possibly due to inherent replication error from DNA polymerase. In the period from 2008, we have seen new DNA recombination methods being published aiming to either (1) simplify Stemmer shuffling, (2) recombine DNA sequences of low sequence identity/homology, or (3) increase the number of crossovers per gene.

QuikChange Shuffling, as implied in the name, is a QuikChange-like protocol (An et al., 2011). Homologous genes, flanked by plasmid-derived DNA sequences, were mixed and sonicated to produce short fragments (50–100 bp). In the subsequent primerless PCR, these fragments were assembled on the basis of sequence homology. The PCR products were then annealed with linearized plasmids together with two pairs of complementary primers (mutagenic primers designed to anneal to a predicted crossover site within the gene) to perform a QuikChange-like amplification of whole plasmid. The method avoided the use of DNaseI that preferentially cleaves DNA at 5′-site of pyrimidines and the need of cloning. The authors reported an average of 2.9 crossovers per gene though the use of specially designed mutagenic primers likely contributed to this increase in crossover number.

Lim et al. reported DNA shuffling using ssDNAs (Lim et al., 2012). Targeted fragments of homologous genes were each amplified by one 5′-phosphorylated primer and one unmodified primer. Upon PCR, 5′-phosphorylated DNA strands were selectively degraded by λ exonuclease. Single-stranded DNA fragments obtained were subsequently assembled *via* overlapping PCR with Klenow fragment (KF). The use of ssDNA and KF eliminated the primerless PCR step that is widely used in DNA shuffling for generating full-length sequence and is often resulting in DNA smearing. However, Lim's method would require crossover sites (and therefore number of crossovers) to be pre-determined.

In the quest to extend the application of DNA recombination to sequences of low homology, Schwaneberg group and Hollfelder group independently reported near homology-independent methods for recombinatorial assembly of DNA fragments (Marienhagen et al., 2012; Villiers et al., 2010). Phosphorothioate-based DNA Recombination (PTRec) is a RE- and ligase-free method (Marienhagen et al., 2012); the principle of which is identical to OmniChange discussed above (Dennig et al., 2011). Briefly, target fragments of family genes and vector were amplified with PS linkage-bearing oligonucleotides. After parallel PCRs, iodine/ethanol treatment produced single-stranded overhangs. Upon hybridization, the resulting DNA constructs were transformed into competent host cells directly. USER Friendly DNA Recombination (USERec), on the other hand, requires multiple enzymes. Target fragments of family genes were amplified with uracil-containing oligonucleotides and *PfuTurbo* C_x Hotstart DNA polymersase. Next, a USER enzyme mix (UDG and endonuclease VIII) was used to produce single-stranded overhangs. Fragments were then joined using T4 DNA ligase and amplified with *PfuTurbo* DNA polymerase for subsequent cloning into an appropriate expression vector. Similar to Lim's method discussed above (Lim et al., 2012), users of PTRec and USERec need to pre-define crossover sites to design corresponding primers. It's worth noting that techniques developed for assembling DNA molecules in a pre-defined order could also be adapted for DNA recombination, examples include Gibson assembly [using an enzyme mixture of 5′-exonuclease, DNA polymerase and DNA ligase; (Gibson et al., 2009)], Class IIS endonuclease-mediated ligation of fragments {*e.g.*, BsaI/Eco31I [5′-GGTCTC(1/3)-3′] in conjunction with T4 DNA ligase (Gao et al., 2013; Yan et al., 2012; Zhou et al., 2013)}, and nicking endonuclease-based DNA fragments assembly [*e.g.*, Nb.BbvCI, Nb.BspQI, Nb.BtsI (Wang et al., 2013b)].

4. Trends in genetic diversity creation

Recent developments in mutant library cloning have been geared towards less DNA manipulation (*e.g.*, gel extraction and purification) and less RE- or ligase-dependence. *In vivo* cloning is stepping into the limelight with its ease and speed since PCR product can be directly transformed together with its recipient vector without further manipulation. The advancement in molecular cloning has undoubtedly transformed the ways random mutagenesis libraries (*e.g.*, epPCR) were prepared with increasing number of methods integrating mutagenesis and cloning into a single step. Recent random mutagenesis methods have been designed to increase the occurrence of transversions and consecutive nucleotide substitutions, which are rare in conventional approaches (*e.g.*, epPCR). Despite being technically more demanding, transposition-based methods that achieve codon substitutions have high potential to explore the entire protein sequence space. Method that results in variable gene length was also developed, leading to a new strategy in directed evolution. In focused mutagenesis, we have noticed the development of multiple QuikChange derivatives. The emphases have been placed on increasing efficiency of QuikChange and simultaneous mutagenesis of multiple sites (distantly or closely spaced sites). *In vivo* methods have also been described, although there remain problems to be tackled (*e.g.*, parental sequence removal and recombination efficiency improvement). For DNA recombination, methods with increased number of crossovers and combinatorial assembly of low homology sequences have been reported using pre-determined crossover sites. Within a relatively short period of time (2008 to date), we have witnessed rigorous development in genetic diversity creation.

A survey of one hundred directed evolution papers published in the past three years (2011–2013) reveals that random mutagenesis of single GOI is the most widely used strategy in genetic diversity creation

(Fig. 3). epPCR with Mutazyme, QuikChange, and Stemmer shuffling remain the most frequently employed method in their respective category, despite the stunning array of methods available. This trend is likely a consequence of five factors: (1) Technical simplicity of epPCR and QuikChange protocol, (2) Commercial availability of some methods (*e.g.*, GeneMorph II and QuikChange Lightning from Agilent, Diversify from Clontech), (3) Proprietary issues complicating wide adoption of some methods in academic research laboratories (*e.g.*, EvoSight, SeSaM, L-Shuffling), (4) Mentality of users (*e.g.*, the attitude of "*don't change the working protocol*") and (5) The lack of a systematic archive of mutagenesis methods and a guideline to facilitate method selection. This survey has motivated us to propose a decision diagram (Fig. 4) to facilitate mutagenesis strategy and method selection when embarking on a new directed evolution project.

5. Current challenges in genetic diversity creation

Although protein engineers have the luxury to choose from a multitude of mutagenesis methods available, challenges remain in genetic diversity creation. The most prominent one that persists today is the "number problem" (Reetz et al., 2008). Using a relatively small protein of 100 residues to illustrate, the corresponding protein sequence space is astronomically huge with 20^{100} constituents ($\sim 1.3 \times 10^{130}$). Sampling this pool of sequences is practically impossible. Fortunately, the need to maintain the structural and functional integrity of an "evolving" protein severely restricts the repertoire of acceptable amino acid substitution (Povolotskaya and Kondrashov, 2010) and the allowable number of mutations per sequence. If we introduce only one amino acid substitution randomly in a 100-residue protein, the number of unique sequences is reduced drastically to 1.9×10^3 [${}_{100}C_1 \times (20 - 1)$; excluding wildtype sequence]. If two or three amino acid substitutions are introduced randomly, the numbers of unique sequences are $\sim 2.0 \times 10^6$ [${}_{100}C_2 \times (20 \times 20 - 1)$] and $\sim 1.3 \times 10^9$ [${}_{100}C_3 \times (20 \times 20 \times 20 - 1)$], respectively. These are manageable library sizes if we consider the capacities of current screening strategies in directed evolution [*e.g.*, agar plate screen ($\sim 10^5$), microtiter plate screen ($\sim 10^4$), selection ($\sim 10^9$), cell surface display ($\sim 10^9$) and *in vitro* compartmentalization ($\sim 10^{10}$); (Leemhuis et al., 2009)]. However, we often find wildtype sequences, redundant sequences and truncated sequences in a gene library for various reasons. Low mutation rate of the mutagenesis method employed, single nt substitution in a codon, preference towards transitions, redundancy of the genetic code, mutational hotspots/bias and mutagenesis methods with no removal of parental template sequence could result in high percentage of wildtype or redundant sequences in a gene library (Wong et al., 2007). To account for these contributing factors, the library size to be screened must be considerably larger than the theoretical values calculated above, which in turn exert pressure on screening. While we are limited by the screening capacity, the only solution is to reduce the library size without compromising its quality. As an example, a codon could be saturated by applying an NNN codon degeneracy (N: A/T/G/C; 64 codons for 20 amino acids). To reduce the number of clones to be examined, the same codon could be randomized using an NNK or an NNS codon degeneracy (K: T/G; S: G/C; 32 codons for 20 amino acids). If we utilize an NDT codon degeneracy (D: A/G/T), we can target 12 amino acids (G, V, L, I, F, Y, N, S, C, D, H, R). This is a balanced mix of aliphatic and aromatic, non-polar and polar, as well as negatively charged and positively charged representatives, while excluding chemically/structurally similar amino acids. Reetz and coworkers demonstrated that NDT library was of higher quality compared to NNK library, despite being smaller in size, by evolving an epoxide hydrolase (Reetz et al., 2008). The quality was measured by higher frequency of positive variants and the magnitude of property improvement (Reetz et al., 2008). The concept of creating a "smarter" gene library (good quality yet small in size) should therefore be extended to all mutagenesis methods, though this is not trivial and remains a huge challenge.

The second challenge, somewhat related to the previous, is the development of a rapid mutagenesis method that favours subsets of amino acid substitutions or acceptable amino acid substitutions. The reasons are twofold: (1) to populate property-specific amino acid

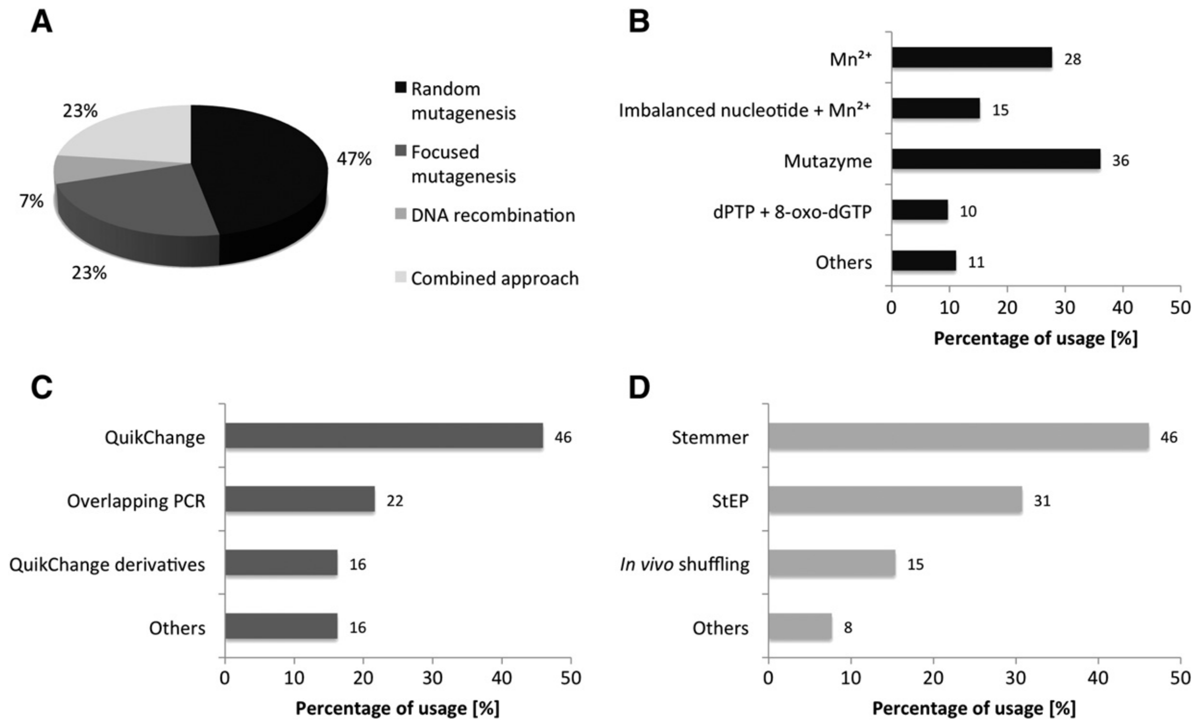

Fig. 3. Trend analysis for genetic diversity creation in directed evolution, based on one hundred randomly selected original articles published between 2011 and 2013. These articles were indexed by PubMed and contained either "*directed evolution*" or "*laboratory evolution*" in the article title. (A) Mutagenesis strategy adopted in these papers, specific method used (B) for random mutagenesis, (C) for focused mutagenesis and (D) for DNA recombination were analysed.

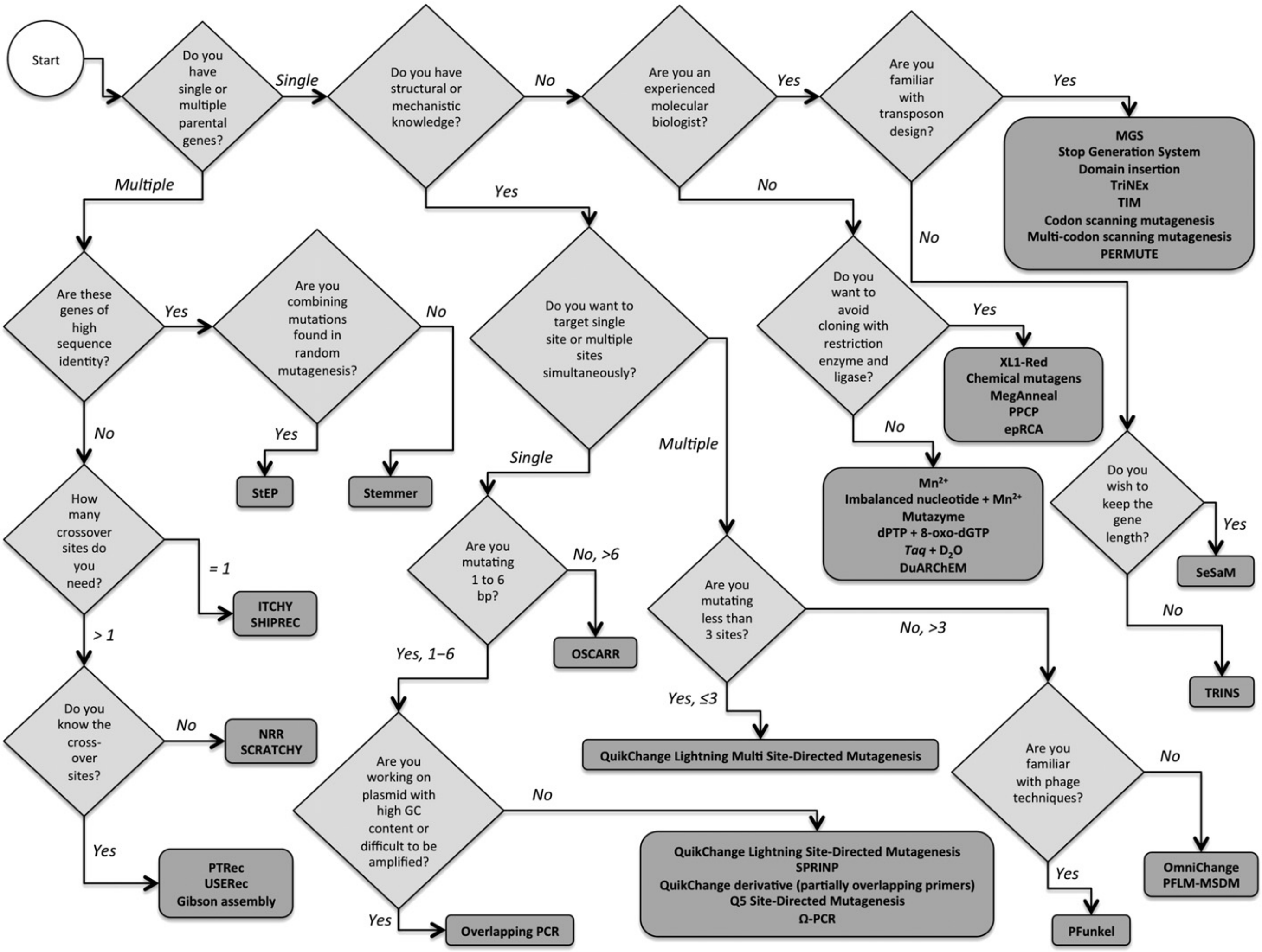

Fig. 4. Decision diagram proposed for mutagenesis strategy and method selection.

substitutions; thereby increasing hit rate, and (2) to reduce the library size and to improve its library quality through eliminating amino acid substitutions that disrupt protein structural and functional integrity. It is widely accepted that amino acid propensity determines protein secondary structure (α-helix, β-sheet or loop). A mutagenesis method that selectively populates certain amino acid substitutions would also shed light on the link between amino acid propensity and protein properties (*e.g.*, thermostability, solvent tolerance). In a recently reported directed evolution of cellobiohydrolase, for example, thermostabilizing mutations found were predominantly S → P substitutions in the loop regions (Wu and Arnold, 2013).

Focused mutagenesis methods that can simultaneously mutate/saturate up to 6 distal sites have been reported. It is however useful to develop methods that can achieve even higher number of codon changes. This proposal is motivated by two lessons that we have learned from protein engineering. First, improved protein variants that accumulated more than 6 mutations after multiple rounds of evolution have been widely reported. Second, the number of amino acids that are involved in substrate binding, protein–protein interaction, protein–nucleic acid interaction or target recognition could well be above 6. Therefore, it would be beneficial to be able to mutate/saturate all these positions simultaneously. OSCARR was developed to mutate a region of up to 16 amino acids (Hidalgo et al., 2008), but there remains a need to develop methods that can simultaneously mutate several regions in a protein to fully explore synergistic effects of mutations in various protein regions.

Numerous DNA recombination methods highlighted in this review achieved increased number of crossovers while recombining non-homologous genes *via* rationally pre-determined crossover sites. Earlier work outside the scope of this review has described method of non-homologous random recombination [NRR (Bittker et al., 2004)]. However, the library created with this method contained many insertions, deletions and rearrangements, which could result in a high number of non-functional protein variants. Whether to apply rationally pre-selected or randomly generated crossover sites likely depends on the structural information available for the proteins of interest. The challenge however remains that we need a robust method which could achieve both high number of random crossovers and low incidence of InDel/rearrangement when recombining genes of low sequence homology.

When developing a mutagenesis method, the difficulty most often encountered is the low number of mutant clones obtained. There are various contributing factors *e.g.* low PCR yield while trying to increase the mutational frequency in a mutagenic PCR (by increasing concentration of $MnCl_2$ or perturbing the balance of nucleotide concentration), low DNA recovery after several DNA purification steps, and poor ligation efficiency. Similarly, increasing the number of simultaneous codon substitutions in focused mutagenesis or increasing the crossover number in DNA recombination often results in lower number of clones. Thus the challenges in the field of creating genetic diversity are multiplex. While seeking methods of greater genetic diversity (quality), one has also to ensure getting sufficient number of clones (quantity) to effectively explore the protein sequence space.

6. Wider applications of mutagenesis

Our discussion on cloning and mutagenesis so far has been limited to phenotypic improvement through accumulation of mutations on a single protein-encoding gene. The application of mutagenesis has, in fact, been extended to multigenic phenotypes or even whole cells. Moreover, mutagenesis is also applied in diverse areas such as viral evolution, metagenomics, and functional study of proteins.

Genome engineering is the art of genome-scale alteration that gives rise to a desired phenotype. Other than allowing us to experimentally probe biological and medical questions (*e.g.*, understanding molecular mechanisms and searching for therapeutic targets in disease treatment), genome engineering has found its biotechnological application in industrial strain improvement. λ Red-mediated recombination, mentioned earlier for cloning mutant libraries, is a widely adopted technique for genome mutagenesis or engineering (Diner et al., 2011). Further, permanent and precise genetic modification of various cell types and organisms could be achieved using customizable molecular scissors that cleave DNA site-specifically resulting in dsDNA breaks (Mussolino and Cathomen, 2013). These damages are subsequently repaired either by homologous recombination, which allows the insertion of new sequences, or by nonhomologous end joining, an error-prone process that can result in gene knockouts. Customizable molecular scissors, reported to date, include ZFNs (also used for molecular cloning), TALENs (transcription activator-like effector nucleases) and RGEN (RNA-guided Endonuclease) (Mussolino and Cathomen, 2013). Phenotype change through temporary reprogramming of gene expression profile has also been demonstrated. Relevant strategies include mutagenesis of global transcription regulators [*e.g.*, cAMP receptor protein (CRP)] (Chong et al., 2013) and creating artificial transcription factors by fusing zinc fingers to either transcription activator or transcription repressor (Park et al., 2003). Church group developed Multiplex Automated Genome Engineering (MAGE) for large-scale programming and evolution of cells (Wang et al., 2009b). Mediated by the bacteriophage λ-Red ssDNA-binding protein β, allelic replacement is achieved in *E. coli* by directing oligonucleotides [designed to target ribosomal binding sites (RBS)] to the lagging strand of the replication fork during DNA replication. Another recombineering-based method, TRackable Multiplex Recombineering (TRMR), was reported by Gill group (Warner et al., 2010). Genetic modifications are created in recombination-proficient *E. coli*, using oligonucleotides designed to target promoter and RBS sequences. These oligonucleotides carry molecular barcodes for subsequent tracking using microarray technique. The development of MAGE and TRMR is stimulated by the decreasing cost of oligonucleotide syntheses and our increasing knowledge of biological systems.

Random mutagenesis and DNA recombination methods have also been applied to viral evolution. The most prominent example is directed evolution of adeno-associated virus (AAV) for enhanced gene delivery, gene targeting, cell-type specificity, capability of crossing blood–brain barrier, and intravitreal transduction (Bartel et al., 2012).

In metagenomics, Truncated Metagenomic Gene-Specific PCR (TMGS-PCR) was developed as a strategy for collecting metagenomic homologous genes for DNA shuffling from environmental samples by Wang et al. (2010). Using lipase as proof of principle, a metagenomic starting gene encoding a protein with lipolytic activity was isolated from functional screening. Based on the sequence of this identified gene, a set of gene-specific primers was designed and used to amplify homologous genes from different environmental samples. The retrieved homologous genes were subjected to conventional DNA shuffling to generate chimeric library.

Oligonucleotide-directed mutagenesis (*e.g.*, QuikChange) is broadly used in alanine-scanning mutagenesis (*i.e.*, systematic amino acid substitution to an alanine), binomial mutagenesis (substitution to an alanine or wildtype amino acid through a degenerate codon) and shotgun mutagenesis (or tetranomial mutagenesis; substitution to an alanine, wildtype amino acids or two other amino acids) (Morrison and Weiss, 2001; Sidhu and Kossiakoff, 2007). These combinatorial libraries with restricted diversity have numerous applications, which include antibody epitope mapping, identifying membrane protein signalling motifs, protein and antibody engineering, optimization of protein stability or expression, mapping of functional domains, and identifying DNA/RNA active elements.

7. Future prospects and conclusion

What does the future hold for genetic diversity creation? We believe a synergistic combination of statistical/computational tools and mutagenesis methods would deliver high quality gene libraries for molecular evolution. Mutagenesis Assistant Program (MAP) and its improved version MAP$^{2.0}$3D, for example, compare amino acid substitution patterns of numerous commonly used random mutagenesis methods upon input of GOI DNA sequence (Verma et al., 2012; Wong et al., 2006a). SCHEMA-guided recombination has met great success in customizing protein properties (Heinzelman et al., 2009, 2010; Romero et al., 2012); PTRec (Marienhagen et al., 2012) and USERec (Villiers et al., 2010) described above are well suited for creating SCHEMA libraries. CASTing is also increasingly used to reduce gene library size (Liang et al., 2007). In addition to these tools that guide our selection of a mutagenesis strategy, programmes are also designed to facilitate molecular biology work. SDM-Assist is a programme that designs primers for site-directed mutagenesis (Karnik et al., 2013) to introduce restriction sites into the mutagenic primers through silent mutations. Therefore, mutant clones could be easily identified through restrictive digestion without DNA sequencing. Tang et al. developed OptiMega, an orthogonal array design by tuning four PCR parameters (concentrations of template, primer, Mg^{2+} and dNTP), to improve methods based on megaprimers (Tang et al., 2013).

Research in nucleic acids has made major advances in the past decades *e.g.* syntheses of nucleotide/nucleobase analogues that expand genetic alphabets beyond A/T/G/C (Henry and Romesberg, 2003). Substrate spectra of DNA polymerases have successfully been engineered to accommodate non-canonical base pairing (Ghadessy et al., 2004; Henry and Romesberg, 2005; Holmberg et al., 2005; Loakes and Holliger, 2009; Loakes et al., 2009). Novel enzymes (*e.g.*, polymerases and endonucleases) are constantly isolated from bioprospecting. In molecular biology, Beer and coworkers demonstrated PCR amplification in under three minutes using near-instantaneous heating and cooling (Wheeler et al., 2011). In the near future, the entire molecular cloning can possibly be performed on a microfluidic chip; the principle of which has been proven by Wang et al. (2011). Advancement in instrument technology coupled with novel enzymes (*e.g.*, faster DNA polymerases) would certainly transform the ways mutant libraries are prepared.

New molecular systems with potential applications in mutagenesis have been described *e.g.* diversity-generating retroelements (DGRs), further expanding our mutagenesis toolbox. DGRs, discovered originally in a *Bordetella* phage BPP-1, are unique family of retroelements that confer selective advantages to their hosts by facilitating localized DNA sequence evolution through a specialized error-prone reverse transcription process (Arambula et al., 2013). Research effort is also channelled into understanding stimulation of spontaneous mutation rate by high transcription rate, a phenomenon widely known as transcription-associated mutation (TAM). In high-transcription yeast strain, for instance, the occurrence of G → T and G → T transversions was elevated by >50 folds (Alexander et al., 2013). Though not a fully developed method for mutant library preparation, TAM could plausibly be an alternative to epPCR for increasing occurrence of transversion mutations.

With research effort now driven towards realizing bioeconomy, directed evolution will continue to play a key role in biotechnology and genetic diversity creation will remain an indispensible tool in many laboratories. The importance of creating high quality mutant libraries is further accentuated by its extended applications in varied research fields. Such widened interest will definitely stimulate continued innovations in methods of genetic diversity creation.

Acknowledgement

We thank The Royal Society (RG110465), ChELSI and EPSRC (EP/E036252/1) for financial support.

References

Abou-Nader M, Benedik MJ. Rapid generation of random mutant libraries. Bioeng Bugs 2010;1:337–40.

Alexander MP, Begins KJ, Crall WC, Holmes MP, Lippert MJ. High levels of transcription stimulate transversions at GC base pairs in yeast. Environ Mol Mutagen 2013;54:44–53.

An Y, Chen L, Sun S, Lv A, Wu W. QuikChange shuffling: a convenient and robust method for site-directed mutagenesis and random recombination of homologous genes. N Biotechnol 2011;28:320–5.

Arambula D, Wong W, Medhekar BA, Guo H, Gingery M, Czornyj E, et al. Surface display of a massively variable lipoprotein by a *Legionella* diversity-generating retroelement. Proc Natl Acad Sci U S A 2013;110:8212–7.

Baldwin AJ, Busse K, Simm AM, Jones DD. Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). Nucleic Acids Res 2008;36:e77.

Bartel MA, Weinstein JR, Schaffer DV. Directed evolution of novel adeno-associated viruses for therapeutic gene delivery. Gene Ther 2012;19:694–700.

Bellini D, Fordham-Skelton AP, Papiz MZ. STRU-cloning: a fast, inexpensive and efficient cloning procedure applicable to both small scale and structural genomics size cloning. Mol Biotechnol 2011;48:30–7.

Bi Y, Qiao X, Hua Z, Zhang L, Liu X, Li L, et al. An asymmetric PCR-based, reliable and rapid single-tube native DNA engineering strategy. BMC Biotechnol 2012;12:39.

Bittker JA, Le BV, Liu JM, Liu DR. Directed evolution of protein enzymes using nonhomologous random recombination. Proc Natl Acad Sci U S A 2004;101:7011–6.

Blanusa M, Schenk A, Sadeghi H, Marienhagen J, Schwaneberg U. Phosphorothioate-based ligase-independent gene cloning (PLICing): an enzyme-free and sequence-independent cloning method. Anal Biochem 2010;406:141–6.

Bloom JD, Arnold FH. In the light of directed evolution: pathways of adaptive protein evolution. Proc Natl Acad Sci U S A 2009;106(Suppl. 1):9995–10000.

Bornscheuer U, Kazlauskas RJ. Survey of protein engineering strategies. Curr Protoc Protein Sci 2011:1–14. [Chapter 26:Unit 7].

Cao W. Endonuclease V: an unusual enzyme for repair of DNA deamination. Cell Mol Life Sci 2012;70:3145–56.

Chen L, Wang F, Wang X, Liu YG. Robust one-tube ohm-PCR strategy accelerates precise sequence modification of plasmids for functional genomics. Plant Cell Physiol 2013;54:634–42.

Chong H, Huang L, Yeow J, Wang I, Zhang H, Song H, et al. Improving ethanol tolerance of *Escherichia coli* by rewiring its global regulator cAMP receptor protein (CRP). PLoS One 2013;8:e57628.

Cobb RE, Si T, Zhao H. Directed evolution: an evolving and enabling synthetic biology tool. Curr Opin Chem Biol 2012;16:285–91.

Daggett KA, Layer M, Cropp TA. A general method for scanning unnatural amino acid mutagenesis. ACS Chem Biol 2009;4:109–13.

Dennig A, Shivange AV, Marienhagen J, Schwaneberg U. OmniChange: the sequence independent method for simultaneous site-saturation of five codons. PLoS One 2011;6:e26222.

Diner EJ, Garza-Sanchez F, Hayes CS. Genome engineering using targeted oligonucleotide libraries and functional selection. Methods Mol Biol 2011;765:71–82.

Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure–function studies. BMC Biotechnol 2009;9:61.

Edwards WR, Busse K, Allemann RK, Jones DD. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res 2008;36:e78.

Erijman A, Dantes A, Bernheim R, Shifman JM, Peleg Y. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. J Struct Biol 2011;175:171–7.

Finney-Manchester SP, Maheshri N. Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM. Nucleic Acids Res 2013;41:e99.

Firnberg E, Ostermeier M. PFunkel: efficient, expansive, user-defined mutagenesis. PLoS One 2012;7:e52031.

Fujii R, Kitaoka M, Hayashi K. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res 2004;32:e145.

Fujii W, Kano K, Sugiura K, Naito K. Repeatable construction method for engineered zinc finger nuclease based on overlap extension PCR and TA-cloning. PLoS One 2013;8:e59801.

Gao X, Yan P, Shen W, Li X, Zhou P, Li Y. Modular construction of plasmids by parallel assembly of linear vector components. Anal Biochem 2013;437:172–7.

Ghadessy FJ, Ramsay N, Boudsocq F, Loakes D, Brown A, Iwai S, et al. Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. Nat Biotechnol 2004;22:755–9.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5.

Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, et al. A family of thermostable fungal cellulases created by structure-guided recombination. Proc Natl Acad Sci U S A 2009;106:5610–5.

Heinzelman P, Komor R, Kanaan A, Romero P, Yu X, Mohler S, et al. Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. Protein Eng Des Sel 2010;23:871–80.

Henry AA, Romesberg FE. Beyond A, C, G and T: augmenting nature's alphabet. Curr Opin Chem Biol 2003;7:727–33.

Henry AA, Romesberg FE. The evolution of DNA polymerases with novel activities. Curr Opin Biotechnol 2005;16:370–7.

Hidalgo A, Schliessmann A, Molina R, Hermoso J, Bornscheuer UT. A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution. Protein Eng Des Sel 2008;21:567–76.

Hoeller BM, Reiter B, Abad S, Graze I, Glieder A. Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). J Microbiol Methods 2008;75:251–7.

Hogrefe HH, Cline J, Youngblood GL, Allen RM. Creating randomized amino acid libraries with the QuikChange multi site-directed mutagenesis kit. Biotechniques 2002;33:1158–60. [62, 64-5].

Holmberg RC, Henry AA, Romesberg FE. Directed evolution of novel polymerases. Biomol Eng 2005;22:39–49.

Huovinen T, Julin M, Sanmark H, Lamminmaki U. Enhanced error-prone RCA mutagenesis by concatemer resolution. Plasmid 2011;66:47–51.

Huovinen T, Brockmann EC, Akter S, Perez-Gamarra S, Yla-Pelto J, Liu Y, et al. Primer extension mutagenesis powered by selective rolling circle amplification. PLoS One 2012;7:e31817.

Jang CW, Magnuson T. A novel selection marker for efficient DNA cloning and recombineering in *E. coli*. PLoS One 2013;8:e57075.

Karnik A, Karnik R, Grefen C. SDM-assist software to design site-directed mutagenesis primers introducing "silent" restriction sites. BMC Bioinformatics 2013;14:105.

Kielkowski P, Macickova-Cahova H, Pohl R, Hocek M. Transient and switchable (triethylsilyl)ethynyl protection of DNA against cleavage by restriction endonucleases. Angew Chem Int Ed Engl 2011;50:8727–30.

Kielkowski P, Brock NL, Dickschat JS, Hocek M. Nucleobase protection strategy for gene cloning and expression. Chembiochem 2013;14:801–4.

Kipnis Y, Dellus-Gur E, Tawfik DS. TRINS: a method for gene modification by randomized tandem repeat insertions. Protein Eng Des Sel 2012;25:437–44.

Le Y, Chen H, Zagursky R, Wu JH, Shao W. Thermostable DNA ligase-mediated PCR production of circular plasmid (PPCP) and its application in directed evolution via in situ error-prone PCR. DNA Res 2013;20:375–82.

Leemhuis H, Kelly RM, Dijkhuizen L. Directed evolution of enzymes: library screening strategies. IUBMB Life 2009;61:222–8.

Liang L, Zhang J, Lin Z. Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing. Microb Cell Fact 2007;6:36.

Lim BN, Choong YS, Ismail A, Glokler J, Konthur Z, Lim TS. Directed evolution of nucleotide-based libraries using lambda exonuclease. Biotechniques 2012;53:357–64.

Liu J, Cropp TA. A method for multi-codon scanning mutagenesis of proteins based on asymmetric transposons. Protein Eng Des Sel 2012;25:67–72.

Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 2008;8:91.

Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem 2010;79:413–44.

Liu P, Hong Y, Lin Y, Fu X, Lin L, Li C, et al. A frequency-controlled random mutagenesis method for GC-rich genes. Anal Biochem 2009;388:356–8.

Loakes D, Holliger P. Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. Chem Commun (Camb) 2009:4619–31.

Loakes D, Gallego J, Pinheiro VB, Kool ET, Holliger P. Evolving a polymerase for hydrophobic base analogues. J Am Chem Soc 2009;131:14827–37.

Marienhagen J, Dennig A, Schwaneberg U. Phosphorothioate-based DNA recombination: an enzyme-free method for the combinatorial assembly of multiple DNA fragments. Biotechniques 2012:1–6.

Martinez R, Jakob F, Tu R, Siegert P, Maurer KH, Schwaneberg U. Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. Biotechnol Bioeng 2013;110:711–20.

Mehta MM, Liu S, Silberg JJ. A transposase strategy for creating libraries of circularly permuted proteins. Nucleic Acids Res 2012;40:e71.

Minamoto T, Wada E, Shimizu I. A new method for random mutagenesis by error-prone polymerase chain reaction using heavy water. J Biotechnol 2012;157:71–4.

Miyazaki K. MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. Methods Enzymol 2011;498:399–406.

Mohan U, Banerjee UC. Molecular evolution of a defined DNA sequence with accumulation of mutations in a single round by a dual approach to random chemical mutagenesis (DuARCheM). Chembiochem 2008;9:2238–43.

Mok WW, Li Y. A highly efficient molecular cloning platform that utilises a small bacterial toxin gene. Chembiochem 2013;14:733–8.

Mori T, Kagatsume I, Shinomiya K, Aoyama Y, Sera T. Sandwiched zinc-finger nucleases harboring a single-chain FokI dimer as a DNA-cleavage domain. Biochem Biophys Res Commun 2009;390:694–7.

Morrison KL, Weiss GA. Combinatorial alanine-scanning. Curr Opin Chem Biol 2001;5:302–7.

Mundhada H, Marienhagen J, Scacioc A, Schenk A, Roccatano D, Schwaneberg U. SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. Chembiochem 2011;12:1595–601.

Mussolino C, Cathomen T. RNA guides genome engineering. Nat Biotechnol 2013;31:208–9.

Pai JC, Entzminger KC, Maynard JA. Restriction enzyme-free construction of random gene mutagenesis libraries in *Escherichia coli*. Anal Biochem 2012;421:640–8.

Park KS, Lee DK, Lee H, Lee Y, Jang YS, Kim YH, et al. Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. Nat Biotechnol 2003;21:1208–14.

Pirakitikulr N, Ostrov N, Peralta-Yahya P, Cornish VW. PCRless library mutagenesis via oligonucleotide recombination in yeast. Protein Sci 2010;19:2336–46.

Povolotskaya IS, Kondrashov FA. Sequence space and the ongoing expansion of the protein universe. Nature 2010;465:922–6.

Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 2009;4:e6441.

Rasila TS, Pajunen MI, Savilahti H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. Anal Biochem 2009;388:71–80.

Reetz MT, Kahakeaw D, Lohmer R. Addressing the numbers problem in directed evolution. Chembiochem 2008;9:1797–804.

Romero PA, Stone E, Lamb C, Chantranupong L, Krause A, Miklos A, et al. SCHEMA designed variants of human arginase I & II reveal sequence elements important to stability and catalysis. ACS Synth Biol 2012;1:221–8.

Sanchis J, Fernandez L, Carballeira JD, Drone J, Gumulya Y, Hobenreich H, et al. Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. Appl Microbiol Biotechnol 2008;81:387–97.

Scholz J, Besir H, Strasser C, Suppmann S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol 2013;13:12.

Shinomiya K, Mori T, Aoyama Y, Sera T. Unidirectional cloning by cleaving heterogeneous sites with a single sandwiched zinc finger nuclease. Biochem Biophys Res Commun 2011;414:733–6.

Shivange AV, Marienhagen J, Mundhada H, Schenk A, Schwaneberg U. Advances in generating functional diversity for directed protein evolution. Curr Opin Chem Biol 2009;13:19–25.

Shivange AV, Serwe A, Dennig A, Roccatano D, Haefner S, Schwaneberg U. Directed evolution of a highly active *Yersinia mollaretii* phytase. Appl Microbiol Biotechnol 2012;95:405–18.

Sidhu SS, Kossiakoff AA. Exploring and designing protein function with restricted diversity. Curr Opin Chem Biol 2007;11:347–54.

Stemmer WP. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A 1994;91:10747–51.

Tang L, Zheng K, Liu Y, Zheng H, Wang H, Song C, et al. Exploring the potential of megaprimer PCR in conjunction with orthogonal array design for mutagenesis library construction. Biotechnol Appl Biochem 2013;60:190–5.

Tseng WC, Lin JW, Wei TY, Fang TY. A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method. Anal Biochem 2008;375:376–8.

Tseng WC, Lin JW, Hung XG, Fang TY. Simultaneous mutations up to six distal sites using a phosphorylation-free and ligase-free polymerase chain reaction-based mutagenesis. Anal Biochem 2010;401:315–7.

Ulrich A, Andersen KR, Schwartz TU. Exponential megapriming PCR (EMP) cloning—seamless DNA insertion into any target plasmid without sequence constraints. PLoS One 2012;7:e53360.

Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol 2010;172:34–44.

Valledor M, Hu Q, Schiller P, Myers RS. Fluorescent protein engineering by in vivo site-directed mutagenesis. IUBMB Life 2012;64:684–9.

Verma R, Schwaneberg U, Roccatano D. MAP2.03D: a sequence/structure based server for protein engineering. ACS Synth Biol. 2012;1:139–50.

Villiers BR, Stein V, Hollfelder F. USER friendly DNA recombination (USERec): a simple and flexible near homology-independent method for gene library construction. Protein Eng Des Sel 2010;23:1–8.

Wang W, Malcolm BA. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. Biotechniques 1999;26:680–2.

Wang BL, Jiao YL, Li XX, Zheng F, Liang H, Sun ZY, et al. A universal method for directional cloning of PCR products based on asymmetric PCR. Biotechnol Appl Biochem 2009a;52:41–4.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 2009b;460:894–8.

Wang Q, Wu H, Wang A, Du P, Pei X, Li H, et al. Prospecting metagenomic enzyme subfamily genes for DNA family shuffling by a novel PCR-based approach. J Biol Chem 2010;285:41509–16.

Wang AB, Cheng CW, Lin IC, Lu FY, Tsai HJ, Lin CC, et al. A novel DNA selection and direct extraction process and its application in DNA recombination. Electrophoresis 2011;32:423–30.

Wang K, Schmied WH, Chin JW. Reprogramming the genetic code: from triplet to quadruplet codes. Angew Chem Int Ed Engl 2012;51:2288–97.

Wang C, Yin X, Kong X, Li W, Ma L, Sun X, et al. A series of TA-based and zero-background vectors for plant functional genomics. PLoS One 2013a;8:e59576.

Wang RY, Shi ZY, Guo YY, Chen JC, Chen GQ. DNA fragments assembly based on nicking enzyme system. PLoS One 2013b;8:e57943.

Wang Z, Wang HY, Feng H. A simple and reproducible method for directed evolution: combination of random mutation with dITP and DNA fragmentation with endonuclease V. Mol Biotechnol 2013c;53:49–54.

Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol 2010;28:856–62.

Wheeler EK, Hara CA, Frank J, Deotte J, Hall SB, Benett W, et al. Under-three minute PCR: probing the limits of fast amplification. Analyst 2011;136:3707–12.

Wong TS, Tee KL, Hauer B, Schwaneberg U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic Acids Res 2004;32:e26.

Wong TS, Tee KL, Hauer B, Schwaneberg U. Sequence saturation mutagenesis with tunable mutation frequencies. Anal Biochem 2005;341:187–9.

Wong TS, Roccatano D, Zacharias M, Schwaneberg U. A statistical analysis of random mutagenesis methods used for directed protein evolution. J Mol Biol 2006a;355:858–71.

Wong TS, Zhurina D, Schwaneberg U. The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen 2006b;9:271–88.

Wong TS, Roccatano D, Schwaneberg U. Challenges of the genetic code for exploring sequence space in directed protein evolution. Biocatal Biotransform 2007;25:229–41.

Wong TS, Roccatano D, Loakes D, Tee KL, Schenk A, Hauer B, et al. Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. Biotechnol J 2008;3:74–82.

Wu I, Arnold FH. Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. Biotechnol Bioeng 2013;110:1874–83.

Wu D, Guo X, Lu J, Sun X, Li F, Chen Y, et al. A rapid and efficient one-step site-directed deletion, insertion, and substitution mutagenesis protocol. Anal Biochem 2013;434:254–8.

Yan P, Gao X, Shen W, Zhou P, Duan J. Parallel assembly for multiple site-directed mutagenesis of plasmids. Anal Biochem 2012;430:65–7.

Yang J, Zhang Z, Zhang XA, Luo Q. A ligation-independent cloning method using nicking DNA endonuclease. Biotechniques 2010;49:817–21.

Yoshida N, Sato M. Plasmid uptake by bacteria: a comparison of methods and efficiencies. Appl Microbiol Biotechnol 2009;83:791–8.

Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 2012;40:e55.

Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res 2004;32:e115.

Zhong X, Zhai C, Yang D, Jiang S, Li Z, Yu X, et al. A single-step mixing cloning method for assembly of lentiviral short hairpin RNA expression vectors for gene silencing. Anal Biochem 2013;438:39–41.

Zhou LB, Lin QQ, Zhang JX, Zhao SJ, Hu ZB. A rapid DNA assembling strategy mediated by direct full-length polymerase chain reaction. Gene 2013;523:122–5.

Zhu D, Zhong X, Tan R, Chen L, Huang G, Li J, et al. High-throughput cloning of human liver complete open reading frames using homologous recombination in *Escherichia coli*. Anal Biochem 2010;397:162–7.