CHAPTER ELEVEN

Overcoming component limitations in synthetic biology through transposon-mediated protein engineering

Joshua T. Atkinson[a], **Bingyan Wu**[b,c], **Laura Segatori**[c,d,e], **Jonathan J. Silberg**[c,d,e,*]

[a]Systems, Synthetic, and Physical Biology Graduate Program, Rice University, Houston, TX, United States
[b]Biochemistry and Cell Biology Graduate Program, Rice University, Houston, TX, United States
[c]Department of Biosciences, Rice University, Houston, TX, United States
[d]Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, United States
[e]Department of Bioengineering, Rice University, Houston, TX, United States
*Corresponding author: e-mail address: joff@rice.edu

Contents

Abstract

Protein fission and fusion can be used to create biomolecules with new structures and functions, including circularly permuted proteins that require post-translational modifications for activity, split protein AND gates that require multiple inputs for activity, and fused domains that function as chemical-dependent protein switches. Herein

Methods in Enzymology, Volume 621
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2019.02.014

© 2019 Elsevier Inc.
All rights reserved.

we describe how transposon mutagenesis can be used for protein design to create libraries of permuted, split, or domain-inserted proteins. When coupled with a functional screen or selection, these approaches can rapidly diversify the topologies and functions of natural proteins and create useful protein components for synthetic biology.

1. Component limitations in synthetic biology

A limited repertoire of well-defined genetically encoded components is available for constructing genetic circuits. These simple parts have enabled the construction of programs with complex dynamic phenotypes, such as toggle switches (Gardner, Cantor, & Collins, 2000), oscillators (Elowitz & Leibler, 2000), logic gates (Gao, Chong, Kim, & Elowitz, 2018; Tamsir, Tabor, & Voigt, 2010), and edge-detectors (Tabor et al., 2009). As synthetic biologists program more complex phenotypes, an increasingly larger set of orthogonal components are required. One-way researchers have diversified the parts list for programming cells is by looking to nature. Libraries of candidate genes identified through metagenomic studies have been synthesized and characterized to identify natural parts with desired functions (van der Helm, Genee, & Sommer, 2018). However, synthetic biology frequently requires genetically encoded components that have not yet been observed in nature, limiting the utility of metagenomic mining as a strategy to identify useful components.

Protein engineering strategies have also been used to overcome component limitations in synthetic biology (Roberta, 2010). Biomolecules with altered functions have been created by screening or selecting libraries of mutated proteins for variants with desired functions. A wide range of methods have been used to create sequence diversity in libraries, such as error-prone PCR (Bloom et al., 2005; Brandsen, Mattheisen, Noel, & Fields, 2018), site-saturation mutagenesis (Öling et al., 2018), homologous recombination (Drummond, Silberg, Meyer, Wilke, & Arnold, 2005), structure-guided recombination (Ho, Adler, Torre, Silberg, & Suh, 2013; Meyer et al., 2003; Otey et al., 2004), and computational design (Dou et al., 2018; Thompson, Bashor, Lim, & Keating, 2012). By altering protein primary structure while conserving protein topology, these approaches have successfully tuned protein substrate specificities, catalytic activities, and stabilities (Arnold, 2009; Dougherty & Arnold, 2009; Renata, Wang, & Arnold, 2015). While these mutagenesis methods have been successfully

used to improve preexisting properties in proteins, they are limited in their potential to create new allosteric properties because they are typically designed to conserve protein topologies.

2. Overcoming component limitations with transposon mutagenesis

In cases where there is a need to control cellular processes on fast timescales, natural proteins can be engineered to display switch-like behaviors through topological modifications. For example, circular permutation has been used to create proteins with altered contact order that are inactive upon expression until they are cut into a pair of fragments by a protease, at which point the resulting fragments associate and cooperatively function (Mitrea, Parsons, & Loh, 2010; Plainkum, Fuchs, Wiyakrutta, & Raines, 2003). Protein fission and fusion can also be used to endow natural proteins with chemical-dependent activities. This latter approach has been achieved by identifying pairs of protein fragments that do not associate and function unless they are fused to a pair of proteins whose interaction is stabilized by chemical binding (Hoff et al., 2009; Pelletier, Campbell-Valois, & Michnick, 1998; Pu, Zinkus-Boltz, & Dickinson, 2017). This approach was recently leveraged to create a split protein sensor that links aggregation of the fused protein to the activity of a transcriptional repressor (Zeng et al., 2018). Protein fragments have also been fused to the termini of ligand-binding domains, such that the activity of the split protein is dependent upon ligand binding. This latter domain insertion approach has been used to achieve post-translational control over fluorescent reporters (Baird, Zacharias, & Tsien, 1999), transcriptional regulation (Younger et al., 2018), antibiotic degradation (Guntas, Mansell, Kim, & Ostermeier, 2005), gene editing (Oakes et al., 2016), metabolic labeling of proteins with non-natural amino acids (Thomas, Pandey, Knudsen, Ball, & Silberg, 2017), and electron transfer (Atkinson et al., 2019).

In vitro transposon mutagenesis (Haapa, Taira, Heikkinen, & Savilahti, 1999) represents a simple strategy to create libraries of vectors that express proteins with altered topologies (Fig. 1). With this approach, a transposase randomly inserts a synthetic transposon into a vector containing a gene of interest, and the product of this reaction is subjected to molecular biology manipulations to create a library of expression vectors with well-defined sequence diversity. This design strategy was initially leveraged to create

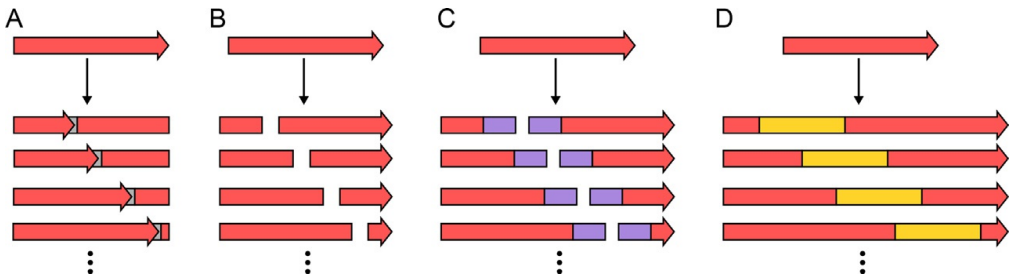

Fig. 1 Topological mutations sampled using transposon mutagenesis. Libraries of vectors can be created that encode different topological variants of a native gene (red), including: (A) circularly permuted genes where the first and last codons of a native gene have been fused through a linker (gray), (B) a pair of gene fragments where two distinct open reading frames are created through fission of a native gene, (C) a pair of gene fragments in which each fragment is fused to genes encoding different proteins (purple), and (D) a pair of gene fragments that are fused to ends of a gene encoding a different protein (yellow), such as a ligand-binding protein.

libraries of vectors that express proteins with pentapeptide insertions (Hoeller, Reiter, Abad, Graze, & Glieder, 2008; Poussu, Vihinen, Paulin, & Savilahti, 2004), hexahistidine insertions (Hoeller et al., 2008; Koerber, Jang, Yu, Kane, & Schaffer, 2007), deletions (Jones, 2005), truncations (Poussu, Jäntti, & Savilahti, 2005), amino acid substitutions (Baldwin, Busse, mm, & Jones, 2008), and non-natural amino acid substitutions (Daggett, Layer, & Cropp, 2009). More recently, transposon mutagenesis has been used to create vector libraries that express proteins with altered topologies, such as randomly split proteins (Segall-Shapiro, Meyer, Ellington, Sontag, & Voigt, 2014; Segall-Shapiro et al., 2011), split proteins fused to pairs of proteins that associate (Mahdavi et al., 2013; Pandey, Nobles, Zechiedrich, Maresso, & Silberg, 2015; Zeng et al., 2018), split proteins fused to the termini of ligand-binding domains (Edwards, Busse, Allemann, & Jones, 2008; Nadler, Morgan, Flamholz, Kortright, & Savage, 2016; Oakes et al., 2016; Thomas et al., 2017), and circularly permuted proteins (Jones et al., 2016; Mehta, Liu, & Silberg, 2012; Pandey et al., 2016). In contrast to early DNAse methods developed for creating libraries of proteins with topological changes (Graf & Schachman, 1996; Hennecke, Sebbel, & Glockshuber, 1999), which are enriched in variants with deletions and insertions of varying lengths, transposon-based methods create well-defined sequence diversity that avoid deletions. This library feature allows for facile analysis using deep mutational scanning (Atkinson, Jones, Zhou, & Silberg, 2018; Higgins & Savage, 2017; Nadler et al., 2016; Oakes et al., 2016).

3. Overview of the library construction workflow

Herein, we describe how transposon mutagenesis can create topological changes in proteins. We highlight how the use of different synthetic transposons in these reactions enables the creation of libraries of vectors that contain genes encoding proteins with different topological changes (Fig. 1), including circularly permuted proteins, split proteins, split proteins fused to other proteins, and split proteins fused to the ends of ligand-binding domains. Finally, we describe how these libraries can be subjected to deep mutational scanning before and after functional analysis to comprehensively characterize protein tolerance to topological changes.

Regardless of the design goal, the first step of all transposon mutagenesis methods involves the insertion of a synthetic transposon into a gene of interest using MuA transposase (Fig. 2). When creating libraries of vectors that encode different topological mutations, slightly different protocols are required. Some methods require that the target gene be encoded within an expression vector, while others require that this gene be circularized. Additionally, some methods yield the desired library in a single step, while others require further manipulation through standard molecular biology

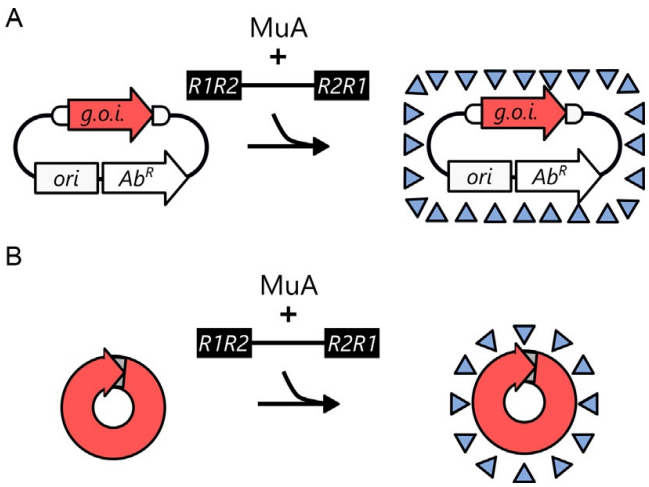

Fig. 2 Transposon insertions occur randomly in the target DNA. Transposition reactions are performed in vitro by mixing an artificial transposon and MuA transposase with (A) a plasmid containing the gene of interest (g.o.i.) or (B) a circularized gene. MuA binds the transposase recognition sequences (R1R2 and R2R1) within the synthetic transposon and randomly inserts it at different locations (triangles) in the target DNA. When performing this reaction with vectors, only transposon insertions within the g.o.i. are desired. All other products of the MuA reaction yield undesired products because they leave the target gene unmodified.

methods, such as replacement of the synthetic transposon with DNA that encodes desired regulatory elements and/or genes that become fused to the target gene. In all cases, careful attention must be paid to maintain stop codons in the correct frame. MuA-mediated DNA insertion results in a 5 base pair duplication at the insertion site, which can affect the relative frame of the open reading frames and the stop codons (Haapa-Paananen, Rita, & Savilahti, 2001).

4. Choosing an artificial transposon

Three classes of synthetic transposons have been reported for building libraries that encode proteins with topological changes. We only discuss previously described transposons below. However, these can be easily modified as long as the transposase recognition sites (R1R2 and R2R1) remain functional within the synthetic transposons. Table 1 provides transposase recognition site sequences that have been shown to function with the transposase MuA.

4.1 Miniature transposons that insert a unique restriction site

Libraries of vectors expressing split proteins and domain-inserted proteins were initially generated using miniature transposons (Edwards et al., 2008; Segall-Shapiro et al., 2011). These synthetic linear DNA sequences contain a selectable cassette, transposase recognition sequences (R1R2 and R2R1), and unique restriction enzyme recognition sequences on the periphery of the R1 sites, such as those cleaved by the Type IIP restriction enzyme NotI (Fig. 3A) or the Golden Gate compatible Type IIS restriction enzyme BsaI (Fig. 3B). When this type of transposon is inserted into a plasmid, it creates an ensemble of vectors that contain the transposon inserted at different sites. Since the transposon is flanked on both sides by the same restriction site, digestion of that site yields sticky ends that can be ligated to compatible synthetic DNA, provided that the plasmid used in the original reaction lacks the restriction sites. A variety of synthetic DNAs have been developed for subcloning in place of the miniature transposon. These include synthetic DNAs that encode regulatory elements required to express split proteins (Fig. 3C), split proteins fused to a pair of proteins (Fig. 3D), split proteins fused to a single protein (Fig. 3E), and split proteins fused to the N- and C-termini of a domain (Fig. 3F).

Table 1 Transposase recognition sequences that can serve as MuA substrates.

Transposon	5′	R1 (5′)	R1 (3′)	3′	Availability
WT Mu	TGAAG	CGGCGCACGAAAAACGCGAAAG	CTTTCGCGTTTTTCGTGCGCCG	CTTCA	–
M1-CamR (NotI)	TGCGG	CCGCGCACGAAAAACGCGAAAG	CTTTCGCGTTTTTCGTGCGCGG	CCGCA	ThermoFisher F-760
Permuteposon P1	TTGAG	CGGCGCACGAAAAACGCGAAAG	CTTTCGCGTTTTTCGTGCGCCG	CTCAA	Addgene #120863
Permuteposon P2	TGAAG	CGGCGCACGAAAAACGCGAAAG	CTTTCGCATTTTGAGTGAGGTA	TATGA	Addgene #120864
Permuteposon P3	TGATT	GATTGAACGAAAAACGCGAAAG	CTTTCGCATTTTGAGTGAATTA	TATGA	Addgene #120865
Splitposon	TGATT	GATTGAACGAAAAACGCGAAAG	CTTTCGCATTTTGAGTGAGGTA	TATGA	Addgene #59947
Mu-BsaI	TG*CAT*	*CG*GAGACCGAAAAACGCGAAAG	CTTTCGCGTTTTTCGGTCTCC*G*	*CGT*CA	Addgene #79769

Synthetic transposons harbor different mutations within the R1 recognition sites. These mutations generate restriction enzyme sites (M1-CamR and Mu-BsaI), stop codons (permuteposon P3 and the splitposon), or ribosomal binding sites (permuteposons P2, P3, and the splitposon). Highlighted nucleotides are mutations relative to wildtype (WT). The Mu transposase recognition sequence is shown as a frame of reference. Bolded nucleotides represent restriction enzyme sites introduced into the recognition sequences through mutation of WT. Underlined nucleotides represent stop (5′) or start codons (3′). Italicized nucleotides represent 4bp overhangs used for downstream Golden Gate ligations (Engler, Kandzia, & Marillonnet, 2008). All of the transposons have wildtype R2 sequences. In the case of the R2 that follows the R1(5′), the sequence is CGTTTCACGATAAATGCGAAAA. In the case of the R2 that precedes the R1(3′), the sequence is TTTTCGCATTTATCGTGAAACG.

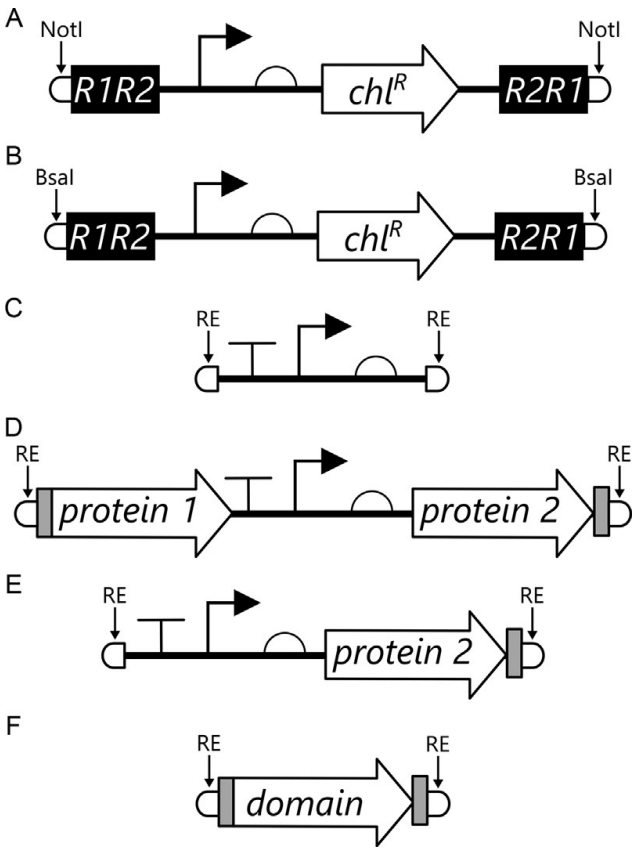

Fig. 3 Transposons that insert unique restriction sites. (A, B) Miniature transposons contain a chloramphenicol resistance cassette (Chl^R) and transposase recognition sequences (R1R2 and R2R1) flanked by a unique restriction site. If this restriction site is absent from the target vector and g.o.i., then the small artificial transposons can be rapidly removed from the product of the insertion reaction through restriction digestion and replaced with synthetic DNA encoding regulatory elements, peptides, or domains. Previously described synthetic DNA include inserts encoding: (C) regulatory elements required to express a split protein without protein fusions (Mahdavi et al., 2013; Segall-Shapiro et al., 2011), (D) a pair of genes with linkers (gray) that become fused to the fragments of the split gene (Pandey et al., 2015; Thomas et al., 2017), (E) a single gene that becomes fused to only one of the fragmented genes (Pandey et al., 2015; Thomas et al., 2017), and (F) a protein domain (Nadler et al., 2016; Oakes et al., 2016; Thomas et al., 2017).

4.2 Splitposons

This synthetic transposon also creates vectors that express different split variants of proteins (Segall-Shapiro et al., 2014). Similar to a miniature transposon, the splitposon contains a selection cassette and transposase recognition sequences (Fig. 4). This synthetic transposon also contains a conditional promoter, a ligand-dependent transcriptional regulator that controls expression from that promoter, and a modified R2R1 that has been mutated to contain a ribosome-binding site (RBS). This hybrid

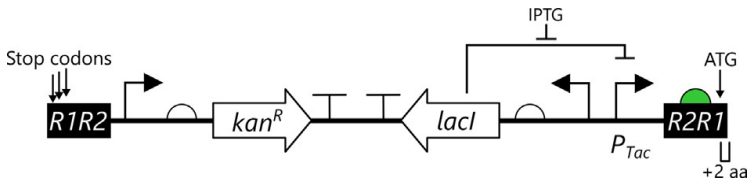

Fig. 4 Splitposon for constructing libraries of fragmented proteins. This synthetic transposon contains a kanamycin resistance cassette (*kan*^{R}), a gene encoding a ligand-dependent transcriptional regulator (*lacI*) that controls transcription from an internal promoter (P_{Tac}), and transposon binding sites (R1R2 and R2R1). The R2R1 site is mutated to contain an RBS (green). This hybrid transposase/ribosome binding site minimizes the number residues added to the N-terminus of the second protein fragment that is under control of P_{Tac} (Segall-Shapiro et al., 2014).

transposase/ribosome binding site, which ultimately controls expression of the C-terminal protein fragment, minimizes the number of residues that are amended to the protein fragment.

4.3 Permuteposons

These synthetic transposons were developed to enable a one-pot synthesis of vectors that express different circularly permuted variants of proteins. When inserted into a circularized gene of interest, permuteposons generate an ensemble of expression vectors since they contain all of the attributes of an expression vector, including a selection cassette, origin of replication, and regulatory elements required to transcribe and translate the circularly permuted genes. Several permuteposons have been developed (Jones et al., 2016; Mehta et al., 2012; Pandey et al., 2016). Some generate a polycistronic mRNA that encodes the selectable marker and the permuted genes (e.g., permuteposons P1, P2, and P3), while others generate a unique mRNA encoding the permuted genes (e.g., P4). The first permuteposon reported (P1) uses an RBS that is located between the selectable cassette and R2R1 to initiate translation of the permuted proteins (Fig. 5A), while the others permuteposons (P2, P3, and P4) use a RBS that is embedded within the transposase binding site (Fig. 5B–D), like the splitposon. P1 adds an 18 amino acid peptide to the N-terminus of all permuted proteins that are expressed (Mehta et al., 2012); the others add 2 amino acids due to the 5 base pair duplication of the MuA insertion site (Jones et al., 2016; Pandey et al., 2016). The advantage of the P1 permuteposon is that it maintains the RBS used to initiate translation in a similar genetic context and yields consistent translation initiation rates across all variants in a library. In contrast, the other permuteposons yield libraries where the circularly permuted protein variants do not experience the same translation initiation rates because of this variability in the genetic context of

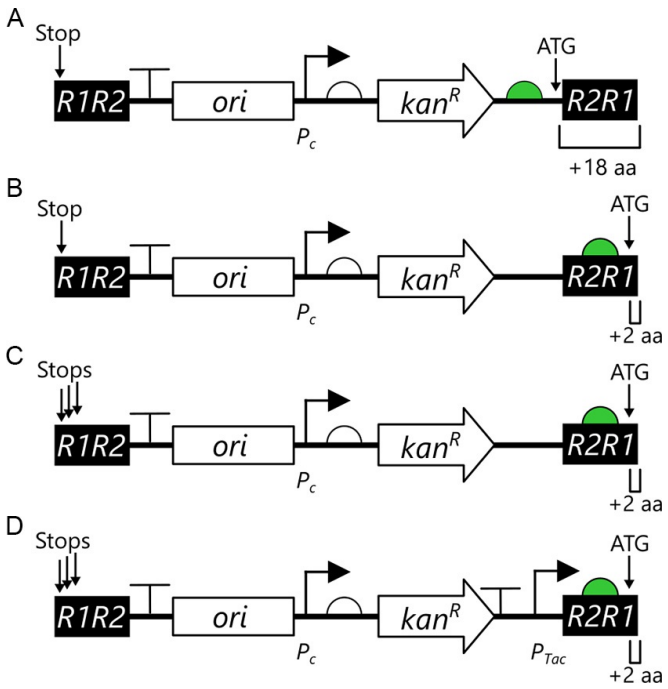

Fig. 5 Permuteposons for building libraries of circularly permuted proteins. (A) Permuteposon P1 contains an origin of replication (ori), a kanamycin resistance cassette (kan^R) that is constitutively expressed using a promoter (P_c), a transcriptional terminator, transposase recognition sequences (R1R2 and R2R1) at the 5′ and 3′ ends (Mehta et al., 2012), and a stop codon to terminate translation of in frame permuted proteins. With P1, the RBS (green) and start codon used to initiate translation of the permuted proteins adds an 18 amino acid peptide to the N-terminus of each permuted variant that is encoded by the R2R1 site. (B) Permuteposon P2 contains a hybrid transposase/ribosome binding site to minimize the number of residues fused to the N-terminus of each permuted variant (Jones et al., 2016). (C) Permuteposon P3 contains a hybrid transposase/ribosome binding site with a different RBS strength from P2. Additionally, P3 includes cascading stop codons that terminate translation occurring in the +1 and −1 reading frames in addition to the in frame stop codon (Jones et al., 2016). (D) Permuteposon P4 uses the same hybrid R2R1 sequence as P2 but includes a terminator and a LacI regulated promoter, P_{Tac}, to control expression of permuted variants (Pandey et al., 2016). P4 also contains cascading stop codons.

the RBS. However, these permuteposons minimize the peptide scar amended to the N-terminus of circularly permuted proteins like splitposons (Jones et al., 2016), which may be desired for certain applications.

5. Library construction

To generate libraries of vectors that express proteins with topological changes, the following materials are required: (1) target DNA encoding the gene of interest, (2) linear synthetic transposon, (3) MuA transposase,

(4) MuA buffer, (5) library-quality electrocompetent *Escherichia coli*, (6) DNA Clean and Concentrator Kit, (7) LB medium and LB-agar plates containing antibiotics, and (8) DNA Miniprep Kit. The first step for generating libraries requires incubating target DNA (500 ng), a synthetic transposon (100 ng), and MuA (0.22 μg; Thermo Fischer F-750) in a 20 μL reaction containing 1× MuA buffer (25 m*M* Tris-HCl pH 8.0 at 25 °C, 10 m*M* $MgCl_2$, 110 m*M* NaCl, 0.05% Triton X-100, and 10% glycerol) as previously outlined for circular permutation (Jones, Atkinson, & Silberg, 2017). This mixture is incubated at 30 °C for 16 h before heat inactivating the MuA at 75°C for 10 min. Following heat inactivation, the reaction is desalted using a DNA Clean and Concentrator Kit. The DNA mixture is then electroporated into *E. coli* cells, such as MegaX DH10B T1R Electrocompetent Cells (Thermo Fisher Scientific), which can yield $>3 \times 10^{10}$ cfu/μg of plasmid DNA. Transformed cells are plated on multiple LB-agar plates that select for the antibiotic resistance encoded by the synthetic transposon, and the plates are incubated overnight at 37 °C. The colony forming units (cfu) on plates are counted, and all colonies are harvested from plates by adding 1 mL LB to the surface of each plate, scraping colonies into a slurry with a sterile spreader, and transferring the cell slurry from all plates into a single 50 mL sterile tube. This cell slurry is mixed, and the "initial vector library" is purified from the cell slurry using a DNA Miniprep Kit (400 μL of cell slurry/DNA Miniprep column). For each design goal listed below, we describe the target DNA, the synthetic transposon, and any additional steps that are required to synthesize a library of expression vectors. Note that if the desired colony counts are not obtained from the initial reaction, then the MuA reaction can be scaled up prior to the electroporation step.

5.1 Split protein libraries created using miniature transposons

When using miniature transposons for library construction (Fig. 6A), the initial vector library created by the MuA reaction requires further manipulation to yield expression vectors. The first manipulation requires digesting the initial library with a restriction enzyme that cuts adjacent to the gene of interest, separating the genes containing inserted transposons from native genes using agarose gel electrophoresis, subcloning the purified gene-transposon hybrids back into fresh vector backbone, transforming the product of the ligation reaction into electrocompetent cells, and plating the transformed cells onto LB-agar containing antibiotic that selects for the vector. After an overnight incubation, the colonies on plates are

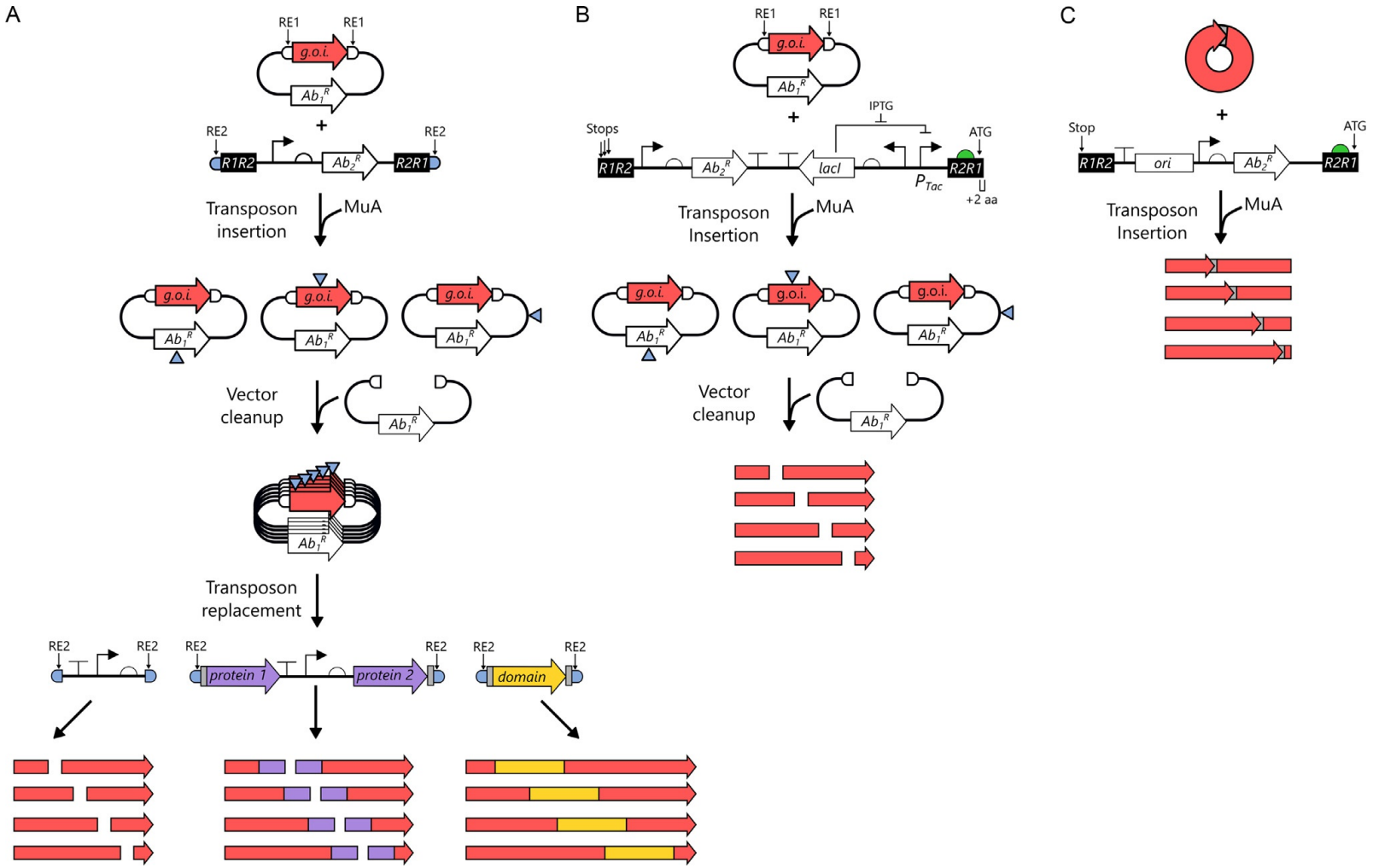

Fig. 6 See legend on opposite page.

harvested as described for the initial vector library, and the resulting vectors are purified from the cell slurry using a DNA Miniprep Kit. This results in a "staged vector library" that has unique restriction sites inserted at different locations within the gene of interest. In contrast to the initial vector library, the staged vector library lacks plasmids containing native genes that are unmodified and vectors having the transposon inserted in the vector backbone.

Once the staged vector library is purified, a variety of split protein libraries can be generated in parallel using this vector ensemble. For all design goals, the synthetic transposons are removed from the staged vector library using restriction enzyme digestion, the linearized vectors lacking the transposons are separated from the transposons using agarose gel electrophoresis, the linear vector ensemble is purified using a DNA Gel Purification Kit, and synthetic DNA is ligated to the vector ensemble to create a final library. The synthetic DNAs for creating split protein libraries are shown in Fig. 3C-E. These yield different kinds of libraries, including libraries of vectors that express split proteins lacking protein fusions (Fig. 3C), split proteins having

Fig. 6 Workflow for creating libraries that express proteins with different classes of topological mutations. (A) When creating vectors that express split proteins or domain-inserted proteins, three steps are required (Pandey et al., 2015; Segall-Shapiro et al., 2011; Thomas et al., 2017). First, transposition reactions are performed in vitro by mixing purified plasmids containing the g.o.i., a miniature transposon, MuA transposase, and MuA buffer. MuA inserts the transposon (triangles) within the gene of interest at different locations and throughout the plasmid. Typically, MuA insertion creates plasmids containing a single inserted transposon. Second, the product of the reaction is digested with a restriction enzyme (RE1) that cuts adjacent to the gene of interest. This yields genes containing or lacking a synthetic transposon, and vector backbone containing or lacking the synthetic transposon. The gene-transposon hybrids are purified and ligated into fresh expression vector to yield a staging library through this vector cleanup process. Third, the synthetic transposons are removed by restriction digestion with a second restriction enzyme (RE2), and a synthetic DNA is subcloned in place of the transposon. By varying the DNA inserted at this step, different topological mutations can be generated. (B) When using splitposons to generate vectors expressing split proteins, only two steps are required for library construction (Segall-Shapiro et al., 2014). First, transposition reactions are performed in vitro by mixing purified plasmids containing a g.o.i., a splitposon, MuA transposase, and MuA buffer. Second, the product of the reaction is digested with a restriction enzyme (RE1) that cuts adjacent to the gene of interest. The purified gene-transposon hybrids are then ligated into fresh expression vector. (C) When using permuteposons to create vectors that express circularly permuted proteins, only one step is required for library construction. Transposition reactions are performed in vitro by mixing a circularized g.o.i., a permuteposon, MuA transposase, and MuA buffer (Atkinson et al., 2018).

the fragments fused to a pair of proteins (Fig. 3D), and split proteins having only the second fragment fused to a protein (Fig. 3E). During the design of these types of synthetic DNA, it is critical to remove stop codons from the genes being fused to the split genes so that open reading frames are generated that express the protein fragments fused to the desired proteins.

5.2 Domain insertion libraries created using miniature transposons

The protocol for creating libraries that express domain inserted proteins is identical to the protocol described for creating split protein libraries with one exception (Fig. 6A). During the last step of library construction, a synthetic DNA encoding a gene that lacks a stop codon (Fig. 3F) is subcloned in place of the synthetic transposon in the staged vector library. It is critical that the synthetic DNA be designed so that the open reading frames encoding the fragments of the split gene are in frame with the inserted gene encoding the domain.

5.3 Split protein libraries created using splitposons

The major benefit of generating libraries with splitposons is that the initial library created by the MuA reaction requires only one processing step to generate the final library (Fig. 6B). With this approach, the initial library contains vectors: (1) having the splitposon inserted within the gene of interest, (2) having the splitposon inserted at other locations, and (3) lacking inserted splitposon. To generate the final library, the gene of interest containing a splitposon inserted in different locations must be subcloned into a fresh vector backbone. This is achieved by digesting the initial vector library with a restriction enzyme that cuts adjacent to the gene of interest, separating this gene-splitposon hybrid from the other digestion products using agarose gel electrophoresis, gel purifying the gene-splitposon hybrids, and subcloning the purified DNA into fresh vector backbone.

5.4 Circularly permuted protein libraries created using permuteposons

Libraries that express circularly permuted variants of a protein can be created in a single step as shown in Fig. 6C. When creating this type of library, the target gene must initially be circularized through ligation prior to mixing with MuA and a permuteposon. The circular gene must lack a stop codon, and the first and last codons in the gene must be linked with DNA consisting of in frame codons that encode for a linker peptide.

6. Sampling considerations when screening versus selecting

Libraries generated by transposon mutagenesis create well-defined sequence diversity that is two times larger than the length of the target gene. This occurs because synthetic transposons are inserted in two orientations. Among the variants created by transposon mutagenesis, only one-sixth of the variants express the desired protein variants with topological changes. This occurs because transposons are randomly inserted in all three frames within the gene of interest and in the two orientations (Fig. 7). The transposon insertion frequency can vary by many orders of magnitude, since MuA does not insert transposons with uniform efficiency across different DNA sequences (Haapa-Paananen et al., 2001). A recent study examined

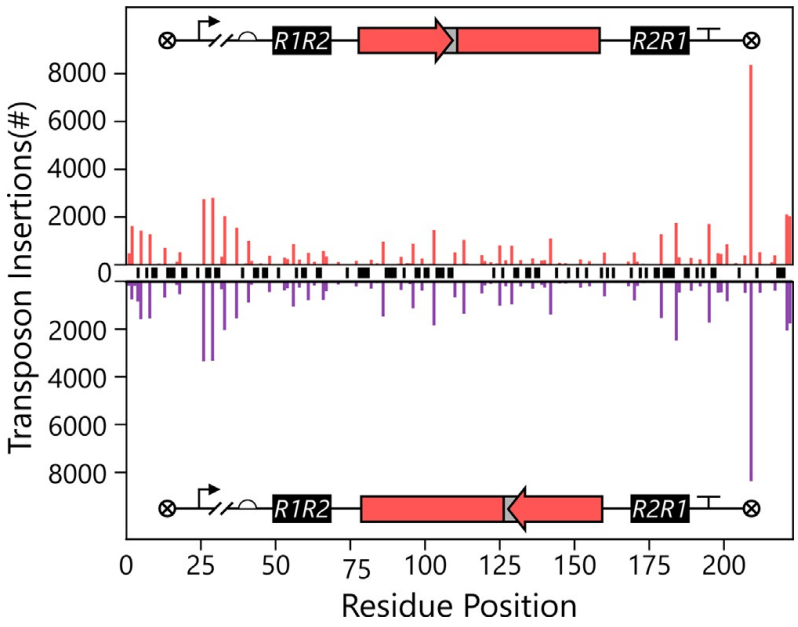

Fig. 7 Positional biases are observed with transposon mutagenesis. Sequencing data from an unselected library of permuted adenylate kinase genes generated by mixing a circular g.o.i. with permuteposon P1 and MuA. When counting insertions in both the forward (red; top) and reverse (purple; bottom) orientations, we observed that a single variant comprised 7% of the total library. Black bars in the center represent the subset of the 223 positions ($n = 64$) that were unsampled in one or both orientations. A total of 114,754 in frame insertions were observed in this library that would theoretically have sampled both orientations in every position of the protein over 100 times. However, due to the bias of MuA, 29% of the positions went unsampled. *Replotted from Atkinson, J. T., Jones, A. M., Zhou, Q., & Silberg, J. J. (2018). Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis.* Nucleic Acids Research, 46*(13), e76.*

the sequence bias when constructing a library of circularly permuted adenylate kinases using a permuteposon (Atkinson et al., 2018). As shown in Fig. 7, this study observed that abundance of individual variants could by vary from 0 to 8377 copies (Atkinson et al., 2018). Even though this library had this level of bias, >70% of the possible vectors that express the different circularly permuted protein variants could be sampled using a functional selection (Atkinson et al., 2018). The bias that occurs with transposase reactions also requires that the initial library construction steps achieve large colony counts from transformations to avoid loss of sequence diversity during the library construction process. The sequence bias observed with libraries is primarily positional. Transposons inserted in the forward and reverse orientations occur with similar frequencies at the same insertion location (Atkinson et al., 2018).

A wide range of methods can be used to mine libraries for functional protein variants (Fig. 8). While low throughput methods such as colony screening on agar plates (or 96-well plates) are limited in their throughput, these have been successfully used to discover proteins with topological mutations that display the desired functional properties (Pandey et al., 2016, 2015). Higher throughput functional assays, such as flow cytometry and bacterial selections, are required when the design goal is to generate a comprehensive profile of a protein's functional tolerance to a topological mutation (Atkinson et al., 2018; Nadler et al., 2016; Oakes et al., 2016).

7. Profiling protein tolerance to topological changes

One of the great advantages of creating libraries through transposon mutagenesis is that this approach generates well-defined sequence diversity that can be analyzed using deep mutational scanning (Atkinson et al., 2018; Nadler et al., 2016; Oakes et al., 2016). Recent studies have illustrated how deep mutational scanning of libraries before and after functional analysis can provide insight into a protein's tolerance to topological changes (Atkinson et al., 2018; Nadler et al., 2016; Oakes et al., 2016). These studies, which are named domain-insertion profiling with DNA sequencing (DIP-seq) and circular permutation profiling with DNA sequencing (CPP-seq), generate profiles that map topological changes that yield functional proteins onto the primary structure of the protein encoded by the target gene. These profiles allow comparison of the sequence abundance of each variant to the

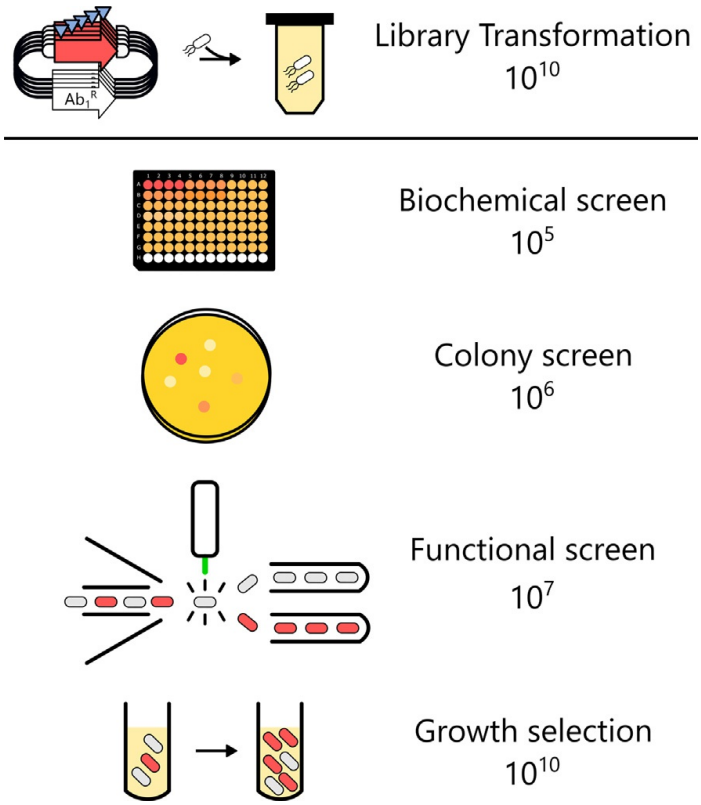

Fig. 8 Strategies for mining libraries for functional protein components. Plasmid libraries created using transposon mutagenesis should be transformed into cells using library-grade competent cells with transformation efficiencies $>10^{10}$ cfu/μg. A variety of strategies can be used to link protein fitness to sequence. Each strategy has a different practical throughput that sets the limit for its ability to comprehensively sample transposon-generated libraries, especially when compounded with the transposon insertion bias (Fig. 7). Biochemical screens of purified proteins are the lowest throughput technique requiring expression, purification, and in vitro activity assay. Screening individual colonies from agar plates (or 96-well plates) using a visual output (e.g., colorimetric substrate or protein) allows for about an order of magnitude increase in throughput. Coupling protein activity to a functional screen in liquid media where cell sorting based on fluorescence can further enhance screening rates. Finally, linking protein activity to cell fitness through a growth selection can greatly increase the number of variants analyzed. These final two strategies are best for generating mutational profiles through deep sequencing as shown in Fig. 9.

abundance of variants that cannot express functional proteins before and after functional enrichment (Fig. 9A). For each variant encoded in the library, the ratio of sequence reads after and before the functional enrichment is calculated and compared to the ratios obtained for vectors that cannot express functional proteins (Fig. 9B). This comparison is used to determine which topological mutants are biologically active. Functional

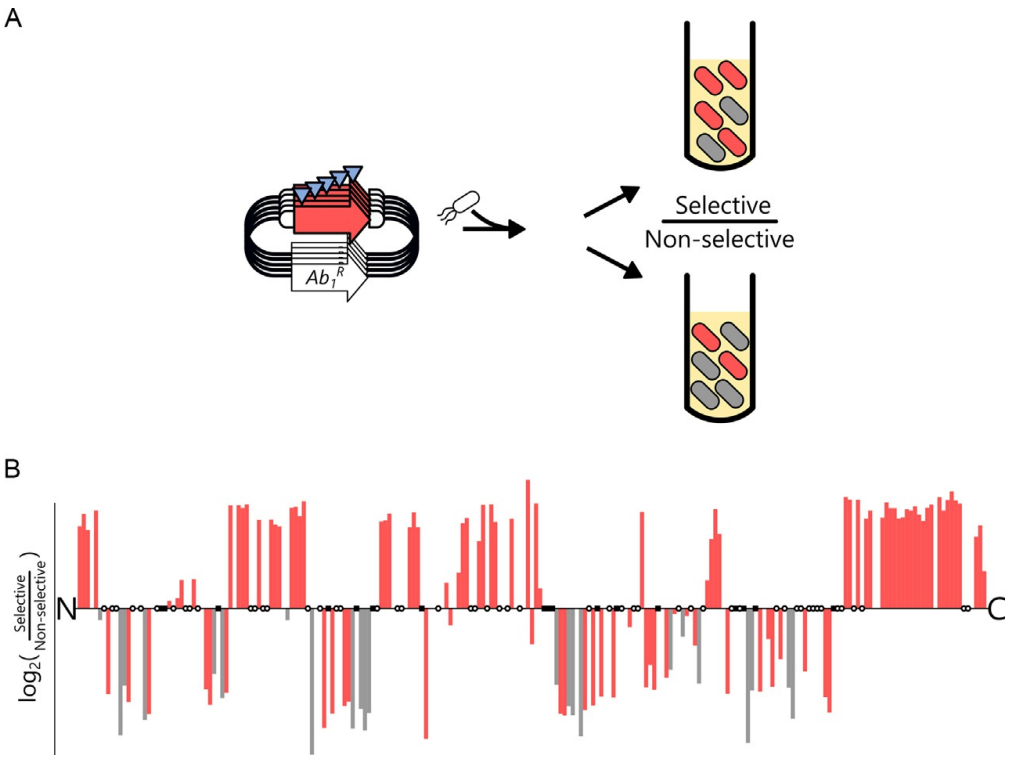

Fig. 9 A profile illustrating protein tolerance to a topological mutation. (A) A plasmid library is subjected to deep mutational scanning where cultures are sequenced before and after subjecting the library to a functional enrichment using a growth selection or functional screen with cell sorting. (B) The $\log_2$-fold ratio of the abundances of each variant observed in the selected versus the unselected conditions is calculated for every native position in the primary structure. These data enable an experimenter to identify positions in the protein primary sequence that retain function following the topological change; only functional variants show significant increases in the selected condition when compared to the unselected condition. Variants having the transposon inserted in the reverse orientation (Fig. 7, bottom panel) are used as a frame of reference for plasmids that cannot express functional variants. CPP-seq data of an adenylate kinase subjected to growth selection after permutation by permuteposon P1. Positions that are significantly enriched relative to vectors that do not express a functional protein are red, while non-significant are in gray. Positions that were not represented in the library are noted with a small circle on the *x* axis. *Replotted from a previously reported data set Atkinson, J. T., Jones, A. M., Zhou, Q., & Silberg, J. J. (2018). Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis.* Nucleic Acids Research, 46*(13), e76.*

mutants are defined as those having sequence ratios that are significantly enriched over the ratios of those variants in the library that cannot express functional proteins. Previous studies have used two different statistical methods for this analysis, including DESeq and Fisher's Exact Test (Atkinson et al., 2018; Nadler et al., 2016; Oakes et al., 2016).

8. Conclusions

Transposon-mediated protein engineering can be used to create combinatorial libraries that are rich in folded and functional proteins with topological mutations, including circularly permuted proteins, split proteins, and domain-inserted proteins. Libraries encoding each of these topological mutations can be rapidly generated with virtually any protein using this methodology. Because transposon mutagenesis libraries are rapid to generate, and they are frequently enriched in proteins with novel switching behaviors, they represent an excellent way to design protein components for synthetic biology applications that require fast, post-translational regulation. Additionally, the libraries generated by transposon mutagenesis contain well-constrained sequence diversity. This feature allows for their analysis through deep mutational scanning before and after functional enrichment and the generation of profiles that comprehensively map a protein's functional tolerance to topological mutations to its primary structure. To date, this approach has only been applied to a handful of proteins. In future studies that apply these methods to a wider range of proteins, we will certainly extend our understanding of the relationship between topological mutations and the evolution of protein allostery.

Acknowledgments

This work was supported by the Office of Naval Research (N00014-17-1-2639), Department of Energy (DE-SC0014462), the National Science Foundation (1805317 and 1615562), and the Welch Foundation (C-1824). J.T.A. was supported by the National Science Foundation Graduate Research Fellowship Program under grant number (R3E821) and Department of Energy Office of Science Graduate Student Research Program.

References

Arnold, F. H. (2009). How proteins adapt: Lessons from directed evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*, 41–46.

Atkinson, J. T., Campbell, I. J., Thomas, E. E., Bonitatibus, S. C., Elliot, S. J., Bennett, G. N., et al. (2019). Metalloprotein switches that display chemical-dependent electron transfer in cells. *Nature Chemical Biology*, *15*(2), 189–195.

Atkinson, J. T., Jones, A. M., Zhou, Q., & Silberg, J. J. (2018). Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis. *Nucleic Acids Research*, *46*(13), e76.

Baird, G., Zacharias, D., & Tsien, R. (1999). Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(20), 11241–11246.

Baldwin, A. J., Busse, K., mm, A., & Jones, D. D. (2008). Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). *Nucleic Acids Research*, *36*(13), e77.

Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, A. D., Adami, C., & Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(3), 606–611.

Brandsen, B. M., Mattheisen, J. M., Noel, T., & Fields, S. (2018). A biosensor strategy for *E. coli* based on ligand-dependent stabilization. *ACS Synthetic Biology*, *7*(9), 1990–1999.

Daggett, K. A., Layer, M., & Cropp, A. T. (2009). A general method for scanning unnatural amino acid mutagenesis. *ACS Chemical Biology*, *4*(2), 109–113.

Dou, J., Vorobieva, A. A., Sheffler, W., Doyle, L. A., Park, H., Bick, M. J., et al. (2018). De novo design of a fluorescence-activating β-barrel. *Nature*, *561*(7724), 485–491.

Dougherty, M. J., & Arnold, F. H. (2009). Directed evolution: New parts and optimized function. *Current Opinion in Biotechnology*, *20*(4), 486–491.

Drummond, A. D., Silberg, J. J., Meyer, M. M., Wilke, C. O., & Arnold, F. H. (2005). On the conservative nature of intragenic recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5380–5385.

Edwards, W. R., Busse, K., Allemann, R. K., & Jones, D. D. (2008). Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Research*, *36*(13), e78.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, *403*(6767), 335.

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, *3*(11), e3647.

Gao, X. J., Chong, L. S., Kim, M. S., & Elowitz, M. B. (2018). Programmable protein circuits in living cells. *Science*, *361*(6408), 1252–1258.

Gardner, T., Cantor, C., & Collins, J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, *403*(6767), 339–342.

Graf, R., & Schachman, H. (1996). Random circular permutation of genes and expressed polypeptide chains: Application of the method to the catalytic chains of aspartate transcarbamoylase. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(21), 11591–11596.

Guntas, G., Mansell, T. J., Kim, J., & Ostermeier, M. (2005). Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(32), 11224–11229.

Haapa, S., Taira, S., Heikkinen, E., & Savilahti, H. (1999). An efficient and accurate integration of mini-Mu transposons in vitro: A general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Research*, *27*(13), 2777–2784.

Haapa-Paananen, S., Rita, H., & Savilahti, H. (2001). DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *The Journal of Biological Chemistry*, *277*(4), 2843–2851.

Hennecke, J., Sebbel, P., & Glockshuber, R. (1999). Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *Journal of Molecular Biology*, *286*(4), 1197–1215.

Higgins, S., & Savage, D. F. (2017). Protein science by DNA sequencing: How advances in molecular biology are accelerating biochemistry. *Biochemistry*, *57*(1), 38–46.

Ho, M. L., Adler, B. A., Torre, M. L., Silberg, J. J., & Suh, J. (2013). SCHEMA computational design of virus capsid chimeras: Calibrating how genome packaging, protection, and transduction correlate with calculated structural disruption. *ACS Synthetic Biology*, *2*(12), 724–733.

Hoeller, B. M., Reiter, B., Abad, S., Graze, I., & Glieder, A. (2008). Random tag insertions by transposon integration mediated mutagenesis (TIM). *Journal of Microbiological Methods*, *75*(2), 251–257.

Hoff, K. G., Culler, S. J., Nguyen, P. Q., McGuire, R. M., Silberg, J. J., & Smolke, C. D. (2009). In vivo fluorescent detection of Fe-S clusters coordinated by human GRX2. *Chemistry & Biology*, *16*(12), 1299–1308.

Jones, D. D. (2005). Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 beta-lactamase to an amino acid deletion. *Nucleic Acids Research*, *33*(9), e80.

Jones, A. M., Atkinson, J. T., & Silberg, J. J. (2017). PERMutation using transposase engineering (PERMUTE): A simple approach for constructing circularly permuted protein libraries. *Methods in Molecular Biology (Clifton, N.J.)*, *1498*, 295–308.

Jones, A. M., Mehta, M. M., Thomas, E. E., Atkinson, J. T., Segall-Shapiro, T. H., Liu, S., et al. (2016). The structure of a thermophilic kinase shapes fitness upon random circular permutation. *ACS Synthetic Biology*, *5*(5), 415–425.

Koerber, J. T., Jang, J.-H., Yu, J. H., Kane, R. S., & Schaffer, D. V. (2007). Engineering adeno-associated virus for one-step purification via immobilized metal affinity chromatography. *Human Gene Therapy*, *18*(4), 367–378.

Mahdavi, A., Segall-Shapiro, T. H., Kou, S., Jindal, G. A., Hoff, K. G., Liu, S., et al. (2013). A genetically encoded AND gate for cell-targeted metabolic labeling of proteins. *Journal of the American Chemical Society*, *135*(8), 2979–2982.

Mehta, M. M., Liu, S., & Silberg, J. J. (2012). A transposase strategy for creating libraries of circularly permuted proteins. *Nucleic Acids Research*, *40*(9), e71.

Meyer, M. M., Silberg, J. J., Voigt, C. A., Endelman, J. B., Mayo, S. L., Wang, Z.-G., et al. (2003). Library analysis of SCHEMA-guided protein recombination. *Protein Science*, *12*(8), 1686–1693.

Mitrea, D. M., Parsons, L. S., & Loh, S. N. (2010). Engineering an artificial zymogen by alternate frame protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 2824–2829.

Nadler, D. C., Morgan, S.-A., Flamholz, A., Kortright, K. E., & Savage, D. F. (2016). Rapid construction of metabolite biosensors using domain-insertion profiling. *Nature Communications*, *7*, 12266.

Oakes, B. L., Nadler, D. C., Flamholz, A., Fellmann, C., Staahl, B. T., Doudna, J. A., et al. (2016). Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nature Biotechnology*, *34*(6), 646–651.

Öling, D., Lawenius, L., Shaw, W., Clark, S., Kettleborough, R., Ellis, T., et al. (2018). Large scale synthetic site saturation GPCR libraries reveal novel mutations that alter glucose signaling. *ACS Synthetic Biology*, *7*(9), 2317–2321.

Otey, C. R., Silberg, J. J., Voigt, C. A., Endelman, J. B., Bandara, G., & Arnold, F. H. (2004). Functional evolution and structural conservation in chimeric cytochromes p450: Calibrating a structure-guided approach. *Chemistry & Biology*, *11*(3), 309–318.

Pandey, N., Kuypers, B. E., Nassif, B., Thomas, E. E., Alnahhas, R. N., Segatori, L., et al. (2016). Tolerance of a knotted near-infrared fluorescent protein to random circular permutation. *Biochemistry*, *55*(27), 3763–3773.

Pandey, N., Nobles, C. L., Zechiedrich, L., Maresso, A. W., & Silberg, J. J. (2015). Combining random gene fission and rational gene fusion to discover near-infrared fluorescent protein fragments that report on protein–protein interactions. *ACS Synthetic Biology*, *4*(5), 615–624.

Pelletier, J., Campbell-Valois, F., & Michnick, S. (1998). Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(21), 12141–12146.

Plainkum, P., Fuchs, S. M., Wiyakrutta, S., & Raines, R. T. (2003). Creation of a zymogen. *Nature Structural Biology*, *10*(2), 115–119.

Poussu, E., Jäntti, J., & Savilahti, H. (2005). A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: Mapping of the Sec1p binding domain in yeast Mso1p by a Mu in vitro transposition-based approach. *Nucleic Acids Research*, *33*(12), e104.

Poussu, E., Vihinen, M., Paulin, L., & Savilahti, H. (2004). Probing the alpha-complementing domain of *E. coli* beta-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu in vitro DNA transposition. *Proteins*, *54*(4), 681–692.

Pu, J., Zinkus-Boltz, J., & Dickinson, B. C. (2017). Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature Chemical Biology*, *13*(4), 432–438.

Renata, H., Wang, J. Z., & Arnold, F. H. (2015). Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angewandte Chemie International Edition*, *54*(11), 3351–3367.

Roberta, K. (2010). Five hard truths for synthetic biology. *Nature*, *463*(7279), 288–290.

Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D., & Voigt, C. A. (2014). A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Molecular Systems Biology*, *10*(7), 742.

Segall-Shapiro, T. H., Nguyen, P. Q., Santos, E. D., Subedi, S., Judd, J., Suh, J., et al. (2011). Mesophilic and hyperthermophilic adenylate kinases differ in their tolerance to random fragmentation. *Journal of Molecular Biology*, *406*(1), 135–148.

Tabor, J. J., lis, H., Simpson, Z., Chevalier, A. A., Levskaya, A., Marcotte, E. M., et al. (2009). A synthetic genetic edge detection program. *Cell*, *137*(7), 1272–1281.

Tamsir, A., Tabor, J. J., & Voigt, C. A. (2010). Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, *469*(7329), 212.

Thomas, E. E., Pandey, N., Knudsen, S., Ball, Z. T., & Silberg, J. J. (2017). Programming post-translational control over the metabolic labeling of cellular proteins with a non-canonical amino acid. *ACS Synthetic Biology*, *6*(8), 1572–1583.

Thompson, K., Bashor, C. J., Lim, W. A., & Keating, A. E. (2012). SYNZIP protein interaction toolbox: In vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synthetic Biology*, *1*(4), 118–129.

Van der Helm, E., Genee, H. J., & Sommer, M. O. (2018). The evolving interface between synthetic biology and functional metagenomics. *Nature Chemical Biology*, *14*(8), 752–759.

Younger, A. K., Su, P. Y., Shepard, A. J., Udani, S. V., Cybulski, T. R., Tyo, K. E., et al. (2018). *Development of novel metabolite-responsive transcription factors via transposon-mediated protein fusion, Protein Engineering, Design, and Selection*, *31*(2), 55–63. https://doi.org/10.1093/protein/gzy001.

Zeng, Y., Jones, A. M., Thomas, E. E., Nassif, B., Silberg, J. J., & Segatori, L. (2018). A split transcriptional repressor that links protein solubility to an orthogonal genetic circuit. *ACS Synthetic Biology*, *7*(9), 2126–2138.