Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Modern methods for laboratory diversification of biomolecules

Sinisa Bratulic and Ahmed H Badran

Genetic variation fuels Darwinian evolution, yet spontaneous mutation rates are maintained at low levels to ensure cellular viability. Low mutation rates preclude the exhaustive exploration of sequence space for protein evolution and genome engineering applications, prompting scientists to develop methods for efficient and targeted diversification of nucleic acid sequences. Directed evolution of biomolecules relies upon the generation of unbiased genetic diversity to discover variants with desirable properties, whereas genome-engineering applications require selective modifications on a genomic scale with minimal off-targets. Here, we review the current toolkit of mutagenesis strategies employed in directed evolution and genome engineering. These state-of-the-art methods enable facile modifications and improvements of single genes, multicomponent pathways, and whole genomes for basic and applied research, while simultaneously paving the way for genome editing therapeutic interventions.

Address

Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Corresponding author: Badran, Ahmed H (ahbadran@broadinstitute.org)

Current Opinion in Chemical Biology 2017, **41**:50–60

This review comes from a themed issue on **Mechanistic biology**

Edited by **Gregory A Weiss**

http://dx.doi.org/10.1016/j.cbpa.2017.10.010

1367-5931/

© 2017 Elsevier Ltd. All rights reserved.

Introduction

Naturally occurring biomolecules are products of evolution in service of the survival of an organism, but they often lack the catalytic efficiency, specificity, or stability necessary for industrial or therapeutic applications. To improve these properties, proteins and other biopolymers are subjected to rounds of directed evolution, a powerful and flexible scheme to systematically endow biomolecules with desirable traits. This approach can employ a variety of mutagenesis strategies to modulate the frequency, distribution, and spectrum of mutations to explore biomolecule sequence landscapes, and applies selections or screens to identify and assess improved variants. The choice of diversification strategy is critical to the success of a directed evolution campaign, as the sequence landscape of a standard protein or biopolymer is typically too vast to be exhaustively searched [1]. Where *a priori* information is limited, unbiased and random *in vitro* [2,3] or *in vivo* [4–7] mutagenesis methods have successfully generated libraries of variants with improved or novel functionalities. Alternatively, bioinformatics, structural, or biochemical information can be leveraged to comprehensively explore a portion of the variant landscape by focusing mutagenesis on functionally relevant sites [8,9]. Finally, newly identified beneficial mutations can be isolated or integrated into single sequences using *in vitro* [10,11] or *in vivo* [12•,13] recombination methods. Recently, laboratory evolution has shifted to techniques that directly couple the diversification and assessment steps, providing the basis for continuous *in vivo* evolution strategies [14] that streamline previously lengthy experiments and minimize the need for human intervention.

Whereas these methods have proven crucial for studying structure-function relationships of single macromolecules [15], functional genomic screening and genome engineering applications typically require unbiased methods for *in situ* genome modification. Early methods to produce strain libraries relied on treatments with chemical mutagens/stressors [16] or transposon mutagenesis [17], and later integrated targeted methods employing homologous recombination capabilities, or *recombineering* (recombination-mediated genetic engineering) [18]. These approaches have been applied in both eukaryotes [19] and prokaryotes [20], and extended to enable *in vivo* continuous genome engineering [21••,22]. The recent discovery of CRISPR-Cas9 systems has reshaped this field, owing to their effectiveness as programmable nucleases or DNA-binding domains, and enabling novel, comparatively facile strategies for targeted diversification in cells [23].

In this review we focus on novel approaches for the diversification of biomolecules and generation of variant libraries, which underlie the application of evolutionary principles in molecular biology research and engineering. We first discuss untargeted mutagenesis methods that are commonly applied in generating diversity for directed protein evolution, and highlight novel methods that have been developed over the past decade. We transition to more targeted mechanisms of creating diversity, and address recent advances in targeted genome modifications with an emphasis on the latest developments in CRISPR-based systems.

Diversification methods for directed protein evolution and functional studies

Random mutagenesis

Natural mutation rates are low ($\sim 10^{-9}$ mutations/nucleotide/generation [24]) and, therefore, inappropriate for diversification of nucleic acid sequences in a laboratory setting. Laboratory evolution of biomolecules critically depends on elevated mutation rates for the discovery of improved or novel activities, mirroring biological principles that exist naturally (e.g. somatic hypermutation is employed to generate substantial antibody diversity [25]). Early protein evolution efforts catalyzed the development of approaches to increase mutation rates in a sequence-independent fashion to facilitate the unbiased construction of large and diverse gene libraries, spearheaded by techniques like error-prone PCR (epPCR). PCR protocols can be modified to reduce the fidelity of the reaction by modulating buffer composition [26] and dNTP ratios [3], introducing nucleoside analogues [2], using proofreading-deficient polymerases [27,28], or treating oligonucleotides with chemical mutagens [29]. While these various approaches increase overall mutation rates, the distribution of specific base changes that are generated can limit the chemical diversity of the resultant libraries. This distribution, called the mutational spectrum, cumulatively describes the efficiency and bias of sequence space exploration by a mutagenesis method. Despite widespread implementation, epPCR suffers from a bias in mutational spectrum (Table 1) to predominately incorporate transitions (A↔G or T↔C), yielding libraries enriched in synonymous mutations or conservative nonsynonymous mutations given the redundancy and assignments of the 64 natural codons. The sequence saturation mutagenesis (SeSaM) [30•] method was developed to specifically address this bias (Table 1), where the promiscuous base-pairing nucleotide inosine is enzymatically incorporated in the variant library and later replaced with canonical nucleotides through standard PCR. A recent improvement, SeSam-Tv-II [31••], increases the likelihood of consecutive mutations, especially double transversions (A/G↔C/T), thereby improving library quality and generating variants that are typically inaccessible by conventional epPCR [31••].

While epPCR-based methods introduce genetic variation primarily through point mutations, insertion and/or deletion (indels) of codons can also have considerable consequences on biomolecule function. These types of variants can be readily accessed using complementary methods such as TRINS [32], which incorporates short tandem repeats generated by rolling circle amplification into a target sequence. The resultant diversity is, however, limited to short sequence duplications rather than truly random insertions, with a significant fraction of the diversified pool encoding frameshifting insertions that can limit downstream discovery efforts (Table 1). It is possible to access in-frame deletions of multiple codons through Mu transposon mutagenesis [33], where the gene of interest bridges a TAT periplasm-directing signal and TEM-1 β-lactamase, ensuring that only an in-frame transposition event creates a functional TAT-β-lactamase product.

Conversely, approaches that do not rely on PCR can simplify library generation by minimizing researcher intervention. In error-prone rolling circle amplification (epRCA) [34], isothermal amplification and mutagenesis are coupled, and RCA-generated libraries can be directly transformed into *Escherichia coli* without further processing by restriction and ligation reactions (Table 1). RCA can also be combined with Kunkel mutagenesis [35] in selective RCA (sRCA) mutagenesis [36]. Plasmids are first produced from an *ung*$^-$ *dut*$^-$ *E. coli* strain to undergo non-specific uridylation (dT→dU), and subsequently amplified by PCR using mutagenic primers. Treatment with uracil-DNA glycosylase (UDG) creates abasic sites in the uracil-containing template, leaving only the mutagenized product for amplification by RCA. The sRCA approach increases effective library sizes and improves the mutagenesis efficiency by eliminating the non-mutated background sequences (Table 1).

Compared to *in vitro* mechanisms that require discreet manipulations to achieve the desired mutagenesis, *in vivo* mutagenesis approaches take inspiration from naturally occurring cellular, error-prone replication machinery. Early attempts at *in vivo* mutagenesis were inspired by SOS response and relied on its components [25], where mutator strains [4] enabled elevated mutagenesis *in vivo*, but to date these approaches have lacked mechanisms to control the resultant mutation rates and spectra (Table 1). In instances where unbiased, whole organism mutagenesis is desirable (e.g. for genome, plasmid, and viral evolution), a mutagenesis plasmid (MP) system encoding inducible dominant mutator alleles can be employed across variable *E. coli* strains. This *in vivo* mutagenesis approach was developed to enable control over a broad dynamic range of mutation rates concomitant with an unbiased mutational spectrum (Figure 1a) [6].

Traditional *in vivo* mutagenesis approaches can also indiscriminately mutagenize the host genome and accessory gene sequences beyond the target locus. This intrinsically limits the mutagenesis rate and may catalyze the appearance of undesirable 'cheaters' in selections or screens. To overcome this limitation, an error-prone Pol I polymerase system was developed to preferentially mutagenize plasmids bearing ColE1 and related origins of replication in *E. coli* over chromosomal sequences [5]. Similarly, a system based on an orthogonal plasmid and an error-prone polymerase pair was recently developed for *in vivo* continuous evolution in yeast [37••]. The orthogonality of mutagenesis in this approach is a significant improvement over previous *in vivo* methods (Table 1),

Download English Version:

https://daneshyari.com/en/article/7693977

Download Persian Version:

https://daneshyari.com/article/7693977

Daneshyari.com