Improved GPCR Engineering with Synthetic Site Saturation Variant Libraries

Find out how synthetic Site Saturation Variant Libraries outperform error-prone PCR in protein engineering

ABSTRACT

Protein engineering and directed evolution applications often use saturation mutagenesis to create a library of variants for screening. However, saturation mutagenesis by error-prone PCR (epPCR) suffers from amplification biases and incomplete access to the codon mutational space. Site Saturation Variant Libraries (SSVLs) offer a synthetic alternative free of the technical limitations of epPCR. Here, the performance of a Twist SSVL was benchmarked against an epPCR library using a glucose activation assay in yeast. Compared to epPCR, the SSVL produced greater variant representation while also simplifying downstream validation steps by providing access to the complete variant diversity.

INTRODUCTION

Saturation mutagenesis is a commonly used protein engineering technique that allows the user to substitute one or more amino acid positions with all the possible codon variants. These variants are then screened to identify mutations that alter the properties or function of the studied protein in a desired manner. Many strategies exist for generating saturation mutagenesis libraries for screening, yet few are able to fully and uniformly represent every possible codon without bias.

Saturation mutagenesis libraries are historically generated through PCR. A strategy called error-prone PCR (epPCR) relies on the error-prone nature of non-proofreading polymerases (e.g., *Taq*) to introduce random mutations in an amplified sequence. However, epPCR is limited by uneven and incomplete variant representation because single-base mutations can only generate a minority of possible codon variants (Kitzman et al., 2015). EpPCR can also inadvertently introduce premature stop codons, making it a non-ideal strategy for large-scale saturation mutagenesis.

PCR can be performed with degenerate primers (i.e., those containing mutant codon sequences) to achieve higher variant representation than epPCR. Even so, multiple factors can limit the success of any PCR approach, including sequence features (e.g., G+C content, melting temperature of the DNA duplex, length), design considerations (i.e., the randomization scheme), premature stop codons, and primer quality (Acevedo-Rocha et al., 2015).

Advances in oligo synthesis technologies have paved the way for fully synthetic saturation mutagenesis libraries. Synthetic libraries have the potential to provide complete or near-complete site representation, unlike PCR-based approaches. Twist's silicon-based platform enables massively high-throughput oligo synthesis with very low error rates (1:2,000). Direct synthesis of each designed mutant results in the production of high-quality saturation mutagenesis libraries with near-complete (approaching 99%) variant representation. The synthetic approach also makes each variant sequence readily available for downstream validation, an advantage not offered by PCR-based strategies.

In collaboration with Twist Bioscience, scientists at AstraZeneca compared the performance of an epPCR library with a synthetic alternative (i.e., a Twist Site-Saturation Variant Library [SSVL]) in a GPCR engineering application (Öling et al., 2018). Using a screening assay to identify novel variants that enhance glucose-sensing by the yeast GPCR Gpr1, this application note highlights the superior variant representation achieved by the Twist SSVL compared to an equivalent epPCR-generated library. The SSVL led to a higher number of enhanced Gpr1 variants than the epPCR library, illustrating the exceptional performance of SSVLs for large-scale saturation mutagenesis applications.

RESULTS AND WORKFLOW

Library Variant Representation

Two site saturation libraries were designed to interrogate the glucose-sensing function of Gpr1: one by epPCR and another by Twist SSVL synthesis. Based on the hypothesis that Gpr1 senses glucose via its extracellular and transmembrane domains, the SSVL targeted amino acids that make up these domains (**Figure 1A**). By contrast, the epPCR library produced random mutations along the entire Gpr1 sequence. Deep sequencing revealed a stark difference in the representation of variants by each approach. Whereas the SSVL possessed a homogeneous distribution of >96% possible variants, the epPCR library contained a heterogeneous distribution of only 35% possible variants (**Figure 1B**).

Library Screening & Performance

Yeast cells express invertase (encoded by the *SUC2* gene) downstream of Gpr1 when exposed to low glucose concentrations. Thus, activation of the *SUC2* promoter was chosen as a surrogate measure of the Gpr1 pathway for screening. Illustrated in **Figure 2**, this pathway proceeds as follows: glucose stimulates Gpr1, causing its G protein α subunit Gpa2 to activate adenylyl cyclase (not pictured). Adenylyl cyclase converts ATP into cyclic-AMP (cAMP). cAMP subsequently activates protein kinase A (PKA), which phosphorylates Rgt1. Phosphorylated Rgt1, which normally represses *SUC2*, dissociates from the *SUC2* promoter, resulting in *SUC2* expression.

Using *SUC2* as a screening readout required a few modifications to yeast. A parallel pathway involving Ras2 also activates adenylyl cyclase and subsequently *SUC2* expression (Gancedo et al, 2015; Xue et al., 1998; **Figure 2**). Thus, Ras2 was deleted to abolish basal activation of the *SUC2* promoter. Wild-type Gpr1 was also deleted to enable reconstitution of the pathway with Gpr1 variants. Yeast bearing both Ras2 and Gpr1 deletions (*gpr1Δras2Δ*) exhibited a significant growth deficit in a leucine reporter assay (see **Materials and Methods**). This strain provided an initial platform for screening the glucose-sensing function of Gpr1 variants.

A super-folder GFP (sfGFP) reporter construct was designed to enable a fluorescent readout of Gpr1 variant function. To that end, sfGFP was inserted in an expression vector downstream of a *SUC2* promoter. A suppressor site (SUC2A) not associated with the Gpr1 pathway was deleted from the *SUC2* promoter in this reporter construct to boost its response to glucose. The reporter construct also contained a constitutive mRuby2 expression cassette to allow for normalization of the fluorescent signal. The resulting sfGFP/mRuby2 reporter plasmid was transformed in the *gpr1Δras2Δ* strain, providing the final strain for screening.

The Gpr1 variant libraries were screened in two steps (**Figure 3A**). Following the transformation of each Gpr1 variant library, ~600 colonies ("mutants") were selected using a leucine reporter to reduce background and to avoid picking variants with an unaltered affinity for glucose (see **Materials and Methods**). The ratio of sfGFP (induced) to mRuby2 (constitutive) was measured for each mutant under basal (no glucose) and induced (0.075% glucose) conditions. **Figure 3B** plots each mutant according to these ratios. Mutants that exhibited >3x the standard deviation of the mean for basal or induced activity were sequenced by Sanger sequencing and validated. Overall, the SSVL generated substantially more unique hits than the epPCR library.

Validation of Gpr1 Variants

Hits were validated by genomically integrating the Gpr1 variant sequence in the sfGFP/mRuby2 reporter strain. As shown in **Figure 3c**, the SSVL identified several mutants that displayed higher basal, but not induced, responsiveness to glucose compared to wild-type Gpr1. These included N116V, I113P, and E138S from extracellular loop 1 (ECL1) as well as W244L and K202G from ECL2. One mutant—A640—showed enhanced responsiveness to glucose under basal and induced conditions (**Figure 4**). This mutant was identified by both ECL3 and the epPCR library.

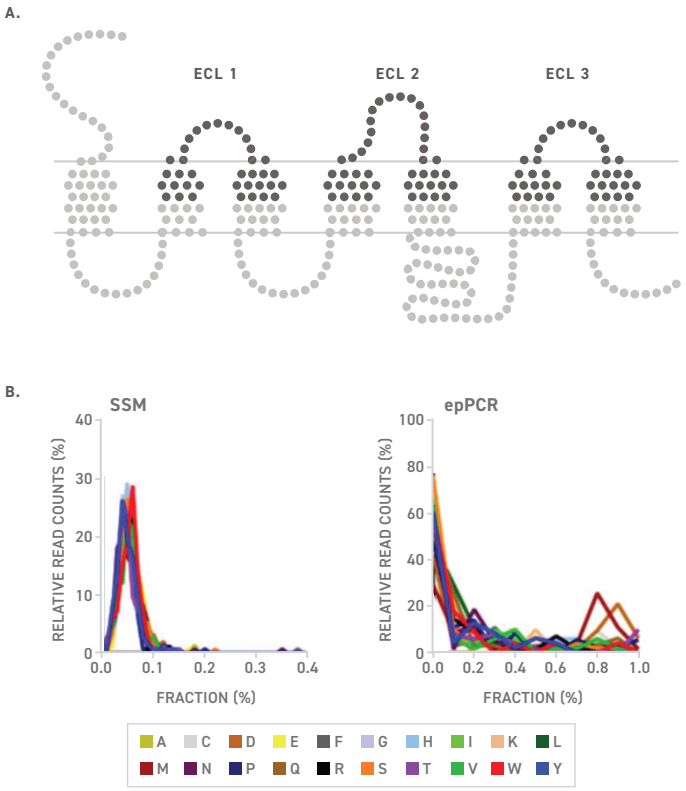

Figure 1. Library design and variant representation. (A) The synthetic SSVL targeted variant positions within the extracellular loops and transmembrane domains of Gpr1 (solid circles). (B) Frequency distribution plots illustrate the distribution of variants in each library.

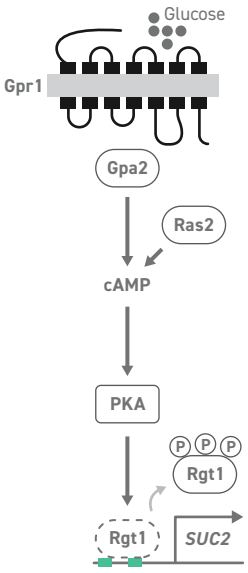

Figure 2. The signaling pathway downstream of Gpr1. SUC2 expression was used as an assay read-out.

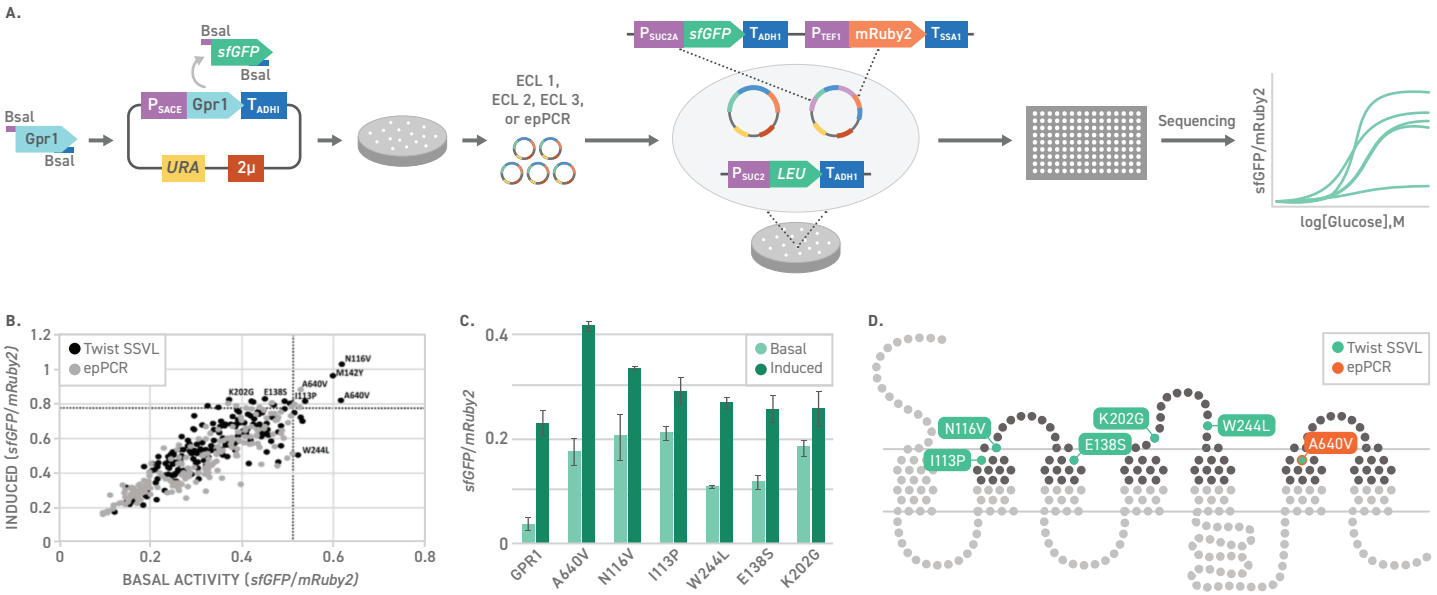

Figure 3. Library performance. (A) Schematic illustrating the screening workflow. Left to right: Library design, Library amplification, the transformation of the library into the reporter yeast strain, screening, data analysis (B) Successful Gpr1 variants identified with the leucine reporter assay were transformed and screened in yeast harboring the sfGFP/mRuby2 reporter plasmid. C) Basal vs. induced sfGFP activity of each identified mutant. The cut-off was set to 3X the standard deviation of wild-type Gpr1. (D) Mutants were identified in each extracellular loop by each library.

CONCLUSIONS

Site saturation mutagenesis is commonly used to screen mutants in protein engineering and directed evolution applications. Using the yeast GPCR Gpr1 as a prototypical example, this application note benchmarked the performance of a synthetic saturation mutagenesis library against one generated by epPCR. Several Gpr1 variants that enhance the receptor's glucose responsiveness were identified by these libraries, including N116V, I113P, E138S, W244L, K202G, and A640V. More specifically, all of these were identified by the SSVL, but only A640V was identified by the epPCR library.

The random nature of epPCR mutagenesis results in variants produced largely by single-base mutations. This restricts the range of codons that are accessible at each position. By contrast, oligo synthesis provides full access to the entire sequence, making all codons possible in an unbiased, uniform distribution. Easy access to the codon mutational space directly impacts the number of possible variants that can be achieved (Kitzman et al, 2015). Indeed, the epPCR and synthetic libraries described here represented 35% and >96% of variants, respectively.

Another key advantage of the SSVL arises during variant validation. With the variant sequences already synthesized, the SSVL allows experimenters to re-clone these variants for validation. Such immediate access is not afforded by epPCR; instead, experimenters must synthesize each variant "hit" for validation. When combined with high variant representation, this time-saving feature makes Twist SSVLs an attractive option for large-scale saturation mutagenesis applications.

PERMISSIONS

REFERENCES

Acevedo-Rocha CG, Reetz MT, & Nov Y (2015) Economical analysis of saturation mutagenesis experiments. Scientific reports, 5, 10654.

Gancedo JM, Flores CL, & Gancedo C (2015) The repressor Rgt1 and the cAMP-dependent protein kinases control the expression of the SUC2 gene in Saccharomyces cerevisiae. Biochimica et biophysica acta, 1850(7), 1362–1367.

Kitzman JO, Starita LM, Lo RS, Fields S, & Shendure J (2015) Massively parallel single-amino-acid mutagenesis. Nature methods, 12(3), 203–206.

Öling D, Lawenius L, Shaw W, Clark S, Kettleborough R, Ellis T, Larsson N, & Wigglesworth M (2018) Large Scale Synthetic Site Saturation GPCR Libraries Reveal Novel Mutations That Alter Glucose Signaling. ACS synthetic biology, 7(9), 2317–2321.

Xue Y, Batlle M, & Hirsch JP (1998) GPR1 encodes a putative G protein-coupled receptor that associates with the Gpa2p Galpha subunit and functions in a Ras-independent pathway. The EMBO journal, 17(7), 1996–2007.