RATIONAL DESIGN OF EXPRESSION VECTORS FOR HIGH-QUALITY BIOLOGICS

Kerry Routenberg Love, Koch Cancer Institute at MIT
kerryluv@mit.edu
Joseph R. Brady, Koch Cancer Institute at MIT
Neil L. Dalvie, Koch Cancer Institute at MIT
J. Christopher Love, Koch Cancer Institute at MIT

Key Words: Functional genomics, transcriptomics, *Pichia pastoris*, transgene cassette, product quality

Commercial proteins (e.g. antibodies, enzymes, vaccine components) for applications from biopharmaceuticals to commodity chemicals require low-cost manufacturing of high-quality products. The engineering of recombinant hosts to achieve large quantities of high-quality heterologous proteins is crucial to both minimizing costs and maximizing safety and efficacy (in the case of biopharmaceuticals). High-quality proteins are properly folded and full-length (intact), with native N-, and C-, termini and bear no significant proteolysis or other degradation (oxidation, deamidation, etc…). As most expression hosts rely on recombinant DNA technology for production of the heterologous protein, the transgene cassette provides an early, and inexpensive, opportunity for optimization of quality and protein titer. Commonly, transgene cassettes include a promoter, a heterologous gene of interest, and terminator for expression of the heterologous gene. A targeting sequence for guided recombination and selective marker for isolation of positive clones are also key elements. In engineering the transgene cassette, factors such as the promoter for heterologous gene expression, target site for transgene integration, sequence for translation initiation, and mRNA codon-optimization of the gene of interest are critical design points for a given protein-expressing strain.

Here, we demonstrate an approach to transgene cassette optimization in the methylotrophic yeast, *Pichia pastoris,* informed by functional genomics. Omics-based techniques such as RNA-Seq, ATAC-Seq and ribosomal foot-printing afford greater upfront understanding for subsequent optimized strain engineering on a product-by-product basis. These types of data are cheap and easy to acquire for yeast and can indicate host- or sequence-derived bottlenecks in transgene transcription, translation and expression. Linking these data to product quality attributes can enlighten the design of the expression vector for fast *in silico* optimization of wide-ranging factors such as gene dosage balance, translation efficiency, and balanced cell kinetics enabling high-quality protein production. Collectively, we show that these tools can enable fast vector design for new heterologous protein-producing strains, including those expressing recombinant vaccines, and robust optimization when engineering higher productivity cell lines.