The monoclonal myth

Thomas C. Scanlon,¹ Heather L. Jewell,¹ Elizabeth C. Gray² and Karl E. Griswold^{1,3,*}

¹Thayer School of Engineering; ²Department of Biological Sciences; and ³Molecular and Cellular Biology Program; Dartmouth College; Hanover, NH USA

Key words: high efficiency transformation, combinatorial library construction, protein engineering, multiple vector transformants, yeast plasmids

Submitted: 02/19/10

Revised: 03/03/10

Accepted: 03/04/10

Previously published online:
www.landesbioscience.com/journals/biobugs/article/11724

*Correspondence to: Karl E. Griswold;
 Email: Karl.E.Griswold@dartmouth.edu

Addendum to: Scanlon TC, Gray EC, Griswold KE. Quantifying and resolving multiple vector transformants in *S. cerevisiae* plasmid libraries. BMC Biotechnol 2009; 9:95; PMID: 19930565; DOI: 10.1186/1472-6750-9-95.

Most researchers confidently assume that transformation of recombinant plasmid libraries into microbial hosts followed by outgrowth of isolated colonies results in a “one cell—one mutant gene—one protein variant” paradigm. Indeed, this assumption is supported by the overwhelming majority of published studies employing bacterial expression hosts. In stark contrast, we recently reported on *Saccharomyces cerevisiae* libraries containing unexpectedly high frequencies of cells harboring heterogeneous mixtures of plasmids, so called Multiple Vector Transformants (MVT). Intriguingly, we observed that yeast MVT persist as a significant proportion of populations for multiple generations. MVT can lead to misidentification of isolated mutants loss of functionally enhanced clones, and unwitting propagation of false positives derived from contaminating control sequences. Such experimental complications can have devastating outcomes in the context of protein engineering by combinatorial library screening. Herein, we demonstrate that the phenomenon of MVT is not restricted to vectors bearing the CEN/ARSH origin of replication, but may be an even greater concern when using high copy 2 μ m plasmids. To mitigate the risks associated with MVT, we have developed an optimized sequencing procedure that facilitates rapid and reliable identification of MVT among clones of interest. In our experience, MVT and their associated risks can be virtually eliminated by employing extended liquid outgrowths of transformed populations and archiving sequence-verified,

monoclonal, mutant genes from cell-templated PCR amplicons.

S. cerevisiae represents a powerful expression and screening platform for protein engineering experiments. With a fully sequenced genome, facile genetics, a multitude of regulatable promoter systems, and a eukaryotic cadre of posttranslational modifying enzymes, there is a good chance that your favorite recombinant protein will express, fold and function in yeast. Additionally, a considerable number of nutritional auxotrophies and counterselection schemes have been combined with yeast's propensity for highly efficient homologous recombination to yield an expansive toolbox for yeast metabolic engineering. Moreover, the capacity to co-opt yeast's dedicated secretion pathways means that many recombinant proteins can be readily secreted to the extracellular space. Driven by its numerous advantages, *S. cerevisiae* is becoming a progressively more popular screening platform for cutting edge biomolecular engineering.

A critical aspect of library-based protein engineering is strict maintenance of genotype-phenotype linkage during high-throughput functional screens. Ideally, screens are performed at the single cell level (alternatively, on a clonal population) and thereby assess the activity of an individual protein variant. Identification and recovery of this protein variant is dependent upon isolation and sequencing of its cognate gene. Although various in vitro methods are available for co-localizing a protein variant and its encoding mutant gene,^{1,2} microbial cells remain the work horses of modern protein engineering. Not

only do living cells provide the full complement of molecular machinery required for transcription, translation and protein folding, but their compartmental nature is ideally suited to rigorous maintenance of the genotype-phenotype link.

In our recent paper entitled, “Quantifying and resolving multiple vector transformants in *S. cerevisiae* plasmid libraries,” we identified a potential liability associated with screening recombinant libraries in yeast. Specifically, we observed a strikingly high frequency of multiple vector transformants (MVT), wherein a single yeast cell is simultaneously transformed with, and temporarily maintains, more than one unique plasmid.³ In the context of protein engineering, cells harboring multiple gene sequences pose a considerable danger that the protein imparting a desired phenotype will be linked with the incorrect mutant gene sequence. Several high profile retractions have been traced to this phenomenon,^{4,5} and MVT may well have played a role in other controversial cases.⁶

Previous work in *E. coli* had examined the issue of bacterial MVT,^{7,8} but we could find no analogous studies of this phenomenon in yeast. Therefore, we set out to quantify the frequency and persistence of MVT in *S. cerevisiae*. Employing low-copy number, centromeric CEN/ARSH plasmids with the URA3 selectable marker, we employed state-of-the-art genetic diversification and transformation techniques to generate recombinant gene libraries in a yeast expression platform. To our great surprise, we observed that up to 95% of transformants harbored more than one mutant gene immediately following high efficiency electroporation. Given that CEN/ARSH plasmids are maintained at less than two copies per cell during steady state growth, we were equally surprised to find that individual MVT harbored, on average, four different vector sequences. Our observations tell a distinctly different story than early studies of CEN/ARSH plasmids in *S. cerevisiae*,⁹ and they have profound implications for those considering implementing yeast as a library screening platform.

Presumably, MVT are resolved into clonal populations via plasmid segregation during cell division. To estimate the

timescale of this process, we monitored MVT frequency as a function of outgrowth time in liquid culture. Remarkably, twenty percent of library transformants were maintained as MVT for at least 24 hours, and 35 hours of outgrowth were required to reduce MVT population frequency below 10%. Thus, MVT not only dominate populations immediately following high efficiency transformation, but they also persist as large subpopulations even after extended outgrowth. Thus, it is clear that MVT are more than just a nuisance in the context of large recombinant yeast libraries.

We have recently expanded the scope of these studies to investigate whether plasmid copy number impacts the frequency of MVT. As opposed to CEN/ARSH plasmids, yeast vectors bearing the 2 μ m origin are maintained at 40–100 copies per cell,¹⁰ and as a result, are commonly used in cases where high level overexpression of a protein is desirable. We hypothesized that high copy number plasmids would yield even larger proportions of MVT than our previous studies with low copy vectors. To test this hypothesis, we reconstructed our original libraries by homologous recombination with a high copy expression plasmid based on p426GAL1.¹¹ Briefly, freshly prepared electrocompetent yeast strain BJ5464 (ATCC#: 208288), were combined with 1 μ g of linearized vector backbone and 2.5 μ g of mutagenic PCR amplicon. The mixture was transformed by high efficiency electroporation, yeast were allowed to recover in 1 M sorbitol for 15 minutes, serial dilutions were plated onto selective media, and plates were incubated at 30°C until colonies appeared (2–3 days). High fidelity cell-templated PCR was performed on individual colonies, the PCR amplicons were sequenced, and the resulting chromatograms were analyzed for the presence of MVT. A schematic overview of the process is shown in **Figure 1**, and detailed experimental procedures are provided in ref. 3. Similar to our prior analysis,³ all observed mutations were restricted to codons that had been specifically targeted in the site-directed library. The population frequency of MVT in a 2 μ m plasmid background was estimated to be 0.85, while that in CEN/ARSH plasmid libraries constructed

under identical conditions was previously estimated to be 0.69. It appears then that the high frequency of yeast MVT is not limited to centromeric plasmids, but is a general phenomenon that should be carefully monitored in any large combinatorial yeast library constructed by electroporation and homologous recombination.

Given the liabilities associated with MVT and their frequency in yeast homologous recombination libraries, we considered whether alternative library construction strategies might produce comparable diversity while circumventing the MVT phenomenon. Using the same 2 μ m vector backbone and ISOR library PCR amplicon as that employed in the homologous recombination studies above, we generated a circularized plasmid library by restriction digestion and T4 ligase mediated end-joining. While we were able to efficiently transform this ligated plasmid library into *E. coli* cells, transformations into freshly prepared electrocompetent *S. cerevisiae* invariably yielded unacceptably poor efficiencies. Similarly, transformation of *S. cerevisiae* with intact circular plasmid libraries yielded significantly fewer transformants than comparable electroporations with linear vector and insert. We note that the usual intent of large, combinatorial, gene libraries is coverage of the greatest possible sequence space. Given our observations that ligation product and intact circular plasmid transform with lower efficiency than linearized DNA, it appears that homologous recombination is the most practical means of plasmid library construction in yeast. Thus, protein engineers using *S. cerevisiae* as an expression and screening platform should be prepared to mitigate the risks associated with high frequency occurrence of MVT in homologous recombination libraries.

A particularly troubling pitfall of the yeast MVT phenomenon is that standard protocols for yeast shuttle vector propagation and maintenance will eliminate evidence of the problem without necessarily addressing it directly. For example, following selection of an interesting yeast clone in a high throughput screen, a standard protocol for sequence identification and archiving would entail: (1) isolating yeast plasmid DNA using a colony patch or

an overnight outgrowth in liquid media, (2) transformation of that plasmid prep into an *E. coli* cloning strain, (3) bacterial transformant outgrowth on selective nutrient plates, (4) liquid outgrowth of individual bacterial colonies, (5) purification of bacterial plasmid DNA, and (6) sequencing the isolated plasmid. This process will take 3–5 days, and importantly, because the MVT frequency in *E. coli* may be comparatively low,⁷ plasmid purified from individual *E. coli* clones will almost certainly contain a single gene variant even if a mixture of plasmid sequences was isolated from the original yeast clone. Therefore, positive identification of MVT would require sequencing of plasmid preps from numerous bacterial colonies, and the desired phenotype would have to be re-verified by transformation of the sequenced plasmid back into the yeast host followed by confirmatory functional screening. Ultimately, this represents a time consuming, resource taxing, labor intensive, and entirely inefficient process that could prompt researchers to spurn the otherwise powerful *S. cerevisiae* expression and screening platform.

To enhance the utility of *S. cerevisiae* as a library screening platform, we have developed an alternative and rapid approach for identifying MVT in clones selected for enhanced functionality. We circumvent the exhausting shuttle vector strategy by direct DNA sequencing of PCR amplified mutant genes. While methods for cell-templated PCR of *S. cerevisiae* have been published elsewhere,¹² it is our experience that inefficient yeast cell lysis is a practical impediment to obtaining reproducible results. In our hands, standard protocols produce PCR amplicons in less than 50% of reactions, and following gel extraction, even successful amplifications frequently yield insufficient DNA for sequencing purposes. We advocate a workflow that includes enzymatic digestion of yeast cell walls with chitinase followed by alkaline lysis with NaOH at 99°C for 10 minutes. The resulting cell lysate is used for PCR with a high fidelity polymerase (e.g., Phusion®, Finnzymes, Finland), and the reaction products are extracted from guanosine supplemented agarose gels following electrophoresis. These purified amplicons can be sequenced directly with

Figure 1. Quantifying yeast MVT with 2 μ m plasmid libraries. Freshly prepared electrocompetent *S. cerevisiae* were transformed with 1 μ g of linearized vector backbone and 2.5 μ g of gene library insert, the latter of which bore flanking segments of homology to the vector. Dilutions of the transformation mixture were plated on selective growth media, and plates were incubated at 30°C until spatially isolated, spherical colonies could be identified. Individual colonies were harvested and used as template in optimized yeast colony PCR reactions with high fidelity Phusion® polymerase. Following gel purification, the PCR amplicons were sequenced, and the resulting chromatograms for each individual colony were visually inspected. Monoclonal colonies showed no overlapping peaks (center chromatogram), while MVTs exhibited multiple peaks at one or more bases specifically targeted by combinatorial mutagenesis (left and right chromatograms). Based on an analysis of 20 randomly selected colonies, the MVT population proportion in the library was estimated to be 0.85 by the Adjusted Wald Method.¹³

greater than 90% success rates. We note that even if sequencing indicates a clonal population, reinsertion of this sequenced amplicon back into the parent expression vector may be the safest way to archive clones of particular interest. All of our developmental work has been undertaken with site-directed saturation libraries, and we therefore know precisely where mutations should and should not be appearing in our isolated genes. We emphasize that after sequencing more than 250

amplicons generated by cell-templated PCR with Phusion® polymerase, we have never observed mutations outside of the target sites. This indicates a low risk that our cell-templated PCR strategy will introduce mutational artifacts into genes selected during functional screening.

In summary, MVT are a practical concern for the protein engineer using *S. cerevisiae* as an expression and screening platform. We have shown that, in the context of *S. cerevisiae* library construction

by high efficiency electroporation, one can confidently assume that multiple sequences will transform the majority of individual yeast. We advocate a liquid outgrowth phase in selective media to ensure that MVT are segregated into clonal populations. Specifically, following electroporation of a combinatorial library, yeast should be allowed to recover in 1 M sorbitol for 15–30 minutes at room temperature. At this point, an aliquot of the transformation mixture can be plated on selective growth agar to assess transformation efficiency. We emphasize that the calculated number of transformants will underestimate the true library diversity, as each colony derived from a MVT encodes multiple mutant genes that will eventually resolve into clonal populations. The remaining transformation mixture should be propagated in selective, repressive,

liquid media at 30°C for at least 48 hours before library screening/archiving. Based on our studies, this approach should minimize the frequency of MVT in library populations.

References

1. Griffiths AD, Tawfik DS. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* 2006; 24:395-402.
2. Zahnd C, Amstutz P, Pluckthun A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat Methods* 2007; 4:269-79.
3. Scanlon TC, Gray EC, Griswold KE. Quantifying and resolving multiple vector transformants in *S. cerevisiae* plasmid libraries. *BMC Biotechnol* 2009; 9:95.
4. Altamirano MM, Blackburn JM, Aguayo C, Fersht AR. Retraction. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* 2002; 417:468.
5. Zeytun A, Jeromin A, Scalettar BA, Waldo GS, Bradbury ARM. Retraction: Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies. *Nat Biotech* 2004; 22:601.
6. Dwyer MA, Looger LL, Hellinga HW. Retraction. *Science* 2008; 319:569.
7. Goldsmith M, Kiss C, Bradbury AR, Tawfik DS. Avoiding and controlling double transformation artifacts. *Protein Eng Des Sel* 2007; 20:315-8.
8. Velappan N, Sblattero D, Chasteen L, Pavlik P, Bradbury AR. Plasmid incompatibility: more compatible than previously thought? *Protein Eng Des Sel* 2007; 20:309-13.
9. Futcher B, Carbon J. Toxic effects of excess cloned centromeres. *Mol Cell Biol* 1986; 6:2213-22.
10. Jordan BE, Mount RC, Hadfield C. Determination of plasmid copy number in yeast. *Meth Mol Biol* 1996; 53:193-203.
11. Mumberg D, Muller R, Funk M. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucl Acids Res* 1994; 22:5767-8.
12. Sambrook J. *Molecular Cloning: A laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press 2001.
13. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998; 52:119-26.